

Exhibit 5



EMORY | LAW

Emory Law Journal

Volume 72 | Issue 3

2023

Consumer Uncertainty in Trademark Law: An Experimental Investigation

Barton Beebe

Roy Germano

Christopher Jon Sprigman

Joel H. Steckel

Follow this and additional works at: <https://scholarlycommons.law.emory.edu/elj>

 Part of the Banking and Finance Law Commons, Commercial Law Commons, and the Consumer Protection Law Commons

Recommended Citation

Barton Beebe, Roy Germano, Christopher J. Sprigman & Joel H. Steckel, *Consumer Uncertainty in Trademark Law: An Experimental Investigation*, 72 Emory L. J. 489 (2023).

Available at: <https://scholarlycommons.law.emory.edu/elj/vol72/iss3/1>

This Article is brought to you for free and open access by the Emory Law Journal at Emory Law Scholarly Commons. It has been accepted for inclusion in Emory Law Journal by an authorized editor of Emory Law Scholarly Commons. For more information, please contact law-scholarly-commons@emory.edu.

CONSUMER UNCERTAINTY IN TRADEMARK LAW: AN EXPERIMENTAL INVESTIGATION

Barton Beebe

Roy Germano

Christopher Jon Sprigman

*Joel H. Steckel**

ABSTRACT

Nearly every important issue in trademark litigation turns on the question of what consumers in the marketplace believe to be true. To address this question, litigants frequently present consumer survey evidence, which can play a decisive role in driving the outcomes of trademark disputes. But trademark survey evidence has often proven to be highly controversial, not least because it has sometimes been perceived as open to expert manipulation. In this Article, we identify and present empirical evidence of a fundamental problem with trademark survey evidence: while the leading survey formats in trademark law test for whether consumers hold a particular belief, they do not examine the strength or the varying degrees of certainty with which consumers hold that belief. Yet as the social science literature has long recognized, the strength with which consumers hold particular beliefs shapes their behavior in the marketplace, and thus it should also shape how trademark disputes play out in the courtroom. Through a series of experiments using the three leading trademark survey formats (the so-called Teflon, Eveready, and Squirt formats), we show the remarkable degree to which these formats as conventionally designed overlook—or suppress—crucial information about consumer

* Barton Beebe, John M. Demarais Professor of Intellectual Property Law, New York University School of Law; Roy Germano, Senior Research Scholar, New York University School of Law; Christopher Jon Sprigman, Murray and Kathleen Bring Professor of Law, New York University School of Law; Joel H. Steckel, Professor of Marketing, NYU Stern School of Business. The authors thank Stefan Bechtold, Christopher Buccafusco, Jeanne Fromer, Daniel Gervais, Glynn Lunney, Irina Manta, and Alex Stremitzer, as well as participants in a trademark scholars online workshop hosted by Mark Lemley and Mark McKenna, and participants in faculty workshops at the New York University School of Law, the University of North Carolina School of Law, ETH-Zürich, the University of Zürich, the 2021 Intellectual Property Scholars Conference, the Tri-State Region IP Workshop held by the Engelberg Center on Innovation Law and Policy, and the International Trademark Association’s Scholarship Symposium for helpful comments and conversations.

uncertainty. We further demonstrate how low-cost, easily administered, and relatively simple modifications to these formats can reveal that information.

We explain both the practical and theoretical implications of our findings. As a practical matter, trademark survey evidence that shows only weakly held beliefs (or that does not even test for belief strength) should not, without more, satisfy a litigant's burden of persuasion on the issue addressed by the survey. Furthermore, in line with courts' growing efforts in intellectual property cases to tailor injunctive relief, survey evidence showing only weakly held mistaken beliefs may provide courts with the opportunity to fashion more limited forms of relief short of an outright injunction. As a theoretical matter, trademark survey formats that reveal the true extent of consumer uncertainty in the marketplace may finally force trademark law and policy to confront normative questions it has long left unanswered going to exactly what kind of harm trademark law is meant to prevent.

TABLE OF CONTENTS

INTRODUCTION	491
I. BACKGROUND: FRAMING CONSUMER UNCERTAINTY IN TRADEMARK LAW	497
A. <i>Trademark Law and the Probabilism of Consumer Beliefs</i>	497
B. <i>The Social Science of Consumer Beliefs</i>	501
C. <i>The Treatment of Consumer Uncertainty in the Early History of Trademark Survey Evidence</i>	506
II. EXPERIMENTS ASSESSING THE EFFECT OF CONSUMER UNCERTAINTY ON TRADEMARK SURVEY METHODS	510
A. <i>The Teflon Genericism Survey Format</i>	511
1. <i>The Origins of the Teflon Survey Format</i>	511
2. <i>Measuring Uncertainty in Teflon Surveys</i>	513
a. <i>Alternative Teflon Method A: Following Up on the Three-Answer Forced Choice Question</i>	514
b. <i>Alternative Teflon Method B: Measuring Uncertainty on a Seven-Point Likert Scale</i>	518
c. <i>Alternative Teflon Method C: Measuring Uncertainty Using a Continuous Scale</i>	520
d. <i>Evaluating the Relative Merits of Alternative Teflon Methods A, B, and C</i>	522
B. <i>The Squirt Confusion Survey Format</i>	523
1. <i>The Origins of the Squirt Survey Format</i>	524
2. <i>Measuring Uncertainty in Squirt Surveys</i>	526

2023]	CONSUMER UNCERTAINTY IN TRADEMARK LAW	491
	<i>a. Alternative Squirt Method A: Following Up on the Three-Answer Forced Choice Question</i>	527
	<i>b. Alternative Squirt Method B: Measuring Uncertainty on a Seven-Point Likert Scale</i>	530
	<i>C. The Eveready Confusion Survey Format</i>	532
	<i>1. The Origins of the Eveready Survey Format</i>	532
	<i>2. Measuring Uncertainty in the Eveready Surveys</i>	534
	III. IMPLICATIONS	537
	<i>A. Consumer Uncertainty and the Expansion of Trademark Rights</i>	538
	<i>B. Incorporating Consumer Uncertainty into Trademark Practice</i>	542
	CONCLUSION	546

INTRODUCTION

What do consumers believe?¹ That is the deceptively simple question on which nearly every important issue in trademark litigation turns. For example, the basic test for trademark infringement asks: Is it likely that a substantial proportion of consumers believes that goods bearing the defendant's trademark originate from the plaintiff?² The basic test to determine if an asserted mark is "generic"³ and thus unprotectable asks: Is it likely that most consumers perceive the mark as referring to an entire category of goods rather than to a specific producer of goods within that category?⁴

¹ J. THOMAS McCARTHY, McCARTHY ON TRADEMARKS AND UNFAIR COMPETITION § 32:158 (5th ed. 2020) ("To an extent not true in other fields of law, in trademark and false advertising disputes the perceptions of large groups of ordinary people are key factual issues. Both trademark validity and infringement turn largely on factual issues of customer perception.").

² See, e.g., Int'l Ass'n of Machinists & Aero. Workers v. Winship Green Nursing Ctr., 103 F.3d 196, 201 (1st Cir. 1996) ("[T]he law has long demanded a showing that the allegedly infringing conduct carries with it a likelihood of confounding an appreciable number of reasonably prudent purchasers exercising ordinary care."); Streetwise Maps, Inc. v. VanDam, Inc., 159 F.3d 739, 743 (2d Cir. 1998) ("A probability of confusion may be found when a large number of purchasers likely will be confused as to the source of the goods in question."); Am. Ass'n for Advancement of Sci. v. Hearst Corp., 498 F. Supp. 244, 258 (D.D.C. 1980) (stating that an "appreciable" number is not necessarily a majority).

³ A "generic" term is the name of a product or service, as distinguished from a term that denotes the *source* of a particular product or service. Generic terms cannot function as trademarks. See RESTATEMENT (THIRD) OF UNFAIR COMPETITION § 15 cmt. a (AM. L. INST. 1995) ("Generic designations are not subject to appropriation as trademarks at common law and are ineligible for registration under state and federal trademark registration statutes.").

⁴ Judge Learned Hand's opinion in *Bayer Co. v. United Drug Co.*, 272 F. 505 (S.D.N.Y. 1921), remains a touchstone. In that opinion, Judge Hand found the term "aspirin" to be generic and stated that "[t]he single question, as I view it, in all these cases, is merely one of fact: What do buyers understand by the word for whose use the parties are contending?" *Id.* at 509.

Trademark law focuses so intently on the question of what consumers believe because consumer beliefs—or courts’ estimates of those beliefs—so often determine whether there is a trademark property right, whether that right has been infringed, and what the remedies for that infringement may be. At every stage of trademark litigation, consumer perception is at the center of the inquiry.⁵

To understand what consumers believe, courts in trademark disputes often consider consumer survey evidence.⁶ Litigants hire survey experts to survey a sample of the relevant consumer population and then testify about their findings. These surveys can play decisive roles in the outcomes of trademark disputes. Consider the recent, closely watched case *United States Patent & Trademark Office v. Booking.com B.V.*⁷ There, the Supreme Court held that adding “.com” to a generic word could result in a distinctive, protectable mark if consumers perceived it as indicating the source of products.⁸ The Court’s ruling in the case

⁵ See Barton Beebe, *Search and Persuasion in Trademark Law*, 103 MICH. L. REV. 2020, 2021 (2005) (“The consumer . . . is the measure of all things in trademark law.”).

⁶ See MCCARTHY, *supra* note 1 (“Survey [e]vidence is [r]outinely [r]eceived and [w]eighed by the [c]ourts.”); Schering Corp. v. Pfizer Inc., 189 F.3d 218, 225 (2d Cir. 1999), *as amended on reh’g* (Sept. 29, 1999) (“Surveys are . . . routinely admitted in trademark and false advertising cases”); *see also* Shari Seidman Diamond & David J. Franklyn, *Trademark Surveys: An Undulating Path*, 92 TEX. L. REV. 2029, 2040 (2014) (“Over time, the use of surveys in trademark and deceptive advertising has grown. According to one account, only 18 surveys were offered in reported cases in the fifteen years between 1946 and 1960, growing to 86 surveys between 1961 and 1975 (approximately 6 per year). Between 1976 to 1990, 442 surveys were presented in reported cases (29 per year); between 1991 and 2005, 742 surveys were offered (approximately 49 per year on average); and in the seven years between 2006 and 2012, about 315 surveys appeared in reported cases (approximately 45 per year).” (internal citations omitted)); Irina Manta, *In Search of Validity: A New Model for the Content and Procedural Treatment of Trademark Infringement Surveys*, 24 CARDozo ARTS & ENT. L. REV. 1016, 1031–32 (2007) (“The importance of consumer surveys in the context of trademark litigation cannot be overemphasized[. . . .] Surveys constitute the main tool to measure ‘the mental state of some segment of the consuming public.’” (quoting Jack P. Lipton, *Trademark Litigation: A New Look at the Use of Social Science Evidence*, 29 ARIZ. L. REV. 639, 642 (1987))). Note, however, that surveys may play a larger role in disputes involving high-value marks. Two comprehensive reviews of survey usage in trademark litigation have found that surveys are in fact not used in most trademark cases. See Barton Beebe, *An Empirical Study of the Multifactor Tests for Trademark Infringement*, 94 CALIF. L. REV. 1581, 1641 (2006) (finding that 20% of the 331 trademark opinions studied discussed survey evidence and 10% credited the survey evidence); Robert C. Bird & Joel H. Steckel, *The Role of Consumer Surveys in Trademark Infringement: Empirical Evidence from the Federal Courts*, 14 U. PA. J. BUS. L. 1013, 1035 (2012) (finding that 16.6% of 533 trademark opinions studied discussed survey evidence); *see also* Graeme W. Austin, *Trademarks and the Burdened Imagination*, 69 BROOK. L. REV. 827, 867–69 (2004) (finding that surveys were introduced in 57.4% of trademark infringement cases that went to final judgment). However, Professors Diamond and Franklyn report that survey evidence is used heavily in lawyers’ pretrial assessment and strategic decision-making, and so the studies reporting relatively low rates of usage in trademark litigation likely underreport the effect of surveys in shaping the outcome of trademark disputes. Diamond & Franklyn, *supra*, at 2061–62.

⁷ 140 S. Ct. 2298 (2020).

⁸ *Id.* at 2301–08.

relied heavily—too heavily, Justice Breyer suggested in dissent⁹—on the respondent’s survey showing that 74.8% of the survey’s respondents perceived BOOKING.COM as a brand name.¹⁰

Though its use is widespread, consumer survey evidence has long been a controversial form of proof in trademark litigation, and for good reason. First, it has sometimes been attacked for being susceptible to expert manipulation. Judge Posner repeatedly questioned the value of “tendentious expert testimony”¹¹ offered by trademark survey experts who are “prone to bias”¹² and took pains in his opinions to expose what he called the “tricks of the survey researcher’s black arts.”¹³ Other judges have lamented that “as is so often the case in high-stakes litigation, highly qualified experts have presented dueling [survey] reports that reach significantly different results.”¹⁴

Second, and more importantly for our purposes here, trademark surveys tend to elide complex normative and empirical questions that underlie trademark law and policy. Among those questions is how trademark law should understand *consumer uncertainty*, i.e., the varying degrees of confidence consumers have in their beliefs in the marketplace. As currently constituted, the leading survey formats provide no way for respondents to indicate *the strength with which they hold a particular belief*—what the social science literature calls “belief strength.”¹⁵ While it is true that each format now conventionally provides respondents with the option to respond “Don’t know/No opinion,” decades of experience show that typically only very low proportions of respondents resort to that response.¹⁶ The remaining response options reduce to a starkly binary

⁹ *Id.* at 2313 (Breyer, J., dissenting) (“[S]ome courts and the TTAB have concluded that survey evidence is generally of little value in separating generic from descriptive terms.”).

¹⁰ *Id.* at 2313–14 (“Consider the survey evidence that respondent introduced below. Respondent’s survey showed that 74.8% of participants thought that ‘Booking.com’ is a brand name, whereas 23.8% believed it was a generic name. At the same time, 33% believed that ‘Washingmachine.com’—which does not correspond to any company—is a brand, and 60.8% thought it was generic.”).

¹¹ *Indianapolis Colts, Inc. v. Metro. Balt. Football Club Ltd. P’ship*, 34 F.3d 410, 415 (7th Cir. 1994); *see also* Robert C. Bird & Joel H. Steckel, *supra* note 6, at 1015 & n.10 (“Expert witnesses can testify about their beliefs regarding confusion, but such testimony may not represent the consumer’s state of mind and can devolve into a ‘battle of the experts’ between hired guns paid to support a particular position.” (internal citation omitted)).

¹² *Kraft Foods Grp. Brands LLC v. Cracker Barrel Old Country Store, Inc.*, 735 F.3d 735, 741 (7th Cir. 2013).

¹³ *Indianapolis Colts, Inc.*, 34 F.3d at 416.

¹⁴ *Bd. of Regents of the Univ. of Hous. Sys. ex rel. the Univ. of Hous. Sys. & Its Member Insts. v. Hous. Coll. Of L., Inc.*, 214 F. Supp. 3d 573, 593 (S.D. Tex. 2016).

¹⁵ *See infra* text accompanying notes 56–58.

¹⁶ *See, e.g., Mattel, Inc. v. MCA Records, Inc.*, 28 F. Supp. 2d 1120, 1134 (C.D. Cal. 1998) (recording that 10% of respondents to a confusion survey question responded that they were not sure), *aff’d*, 296 F.3d 894 (9th Cir. 2002); *Nestle Co. v. Chester’s Mkt., Inc.*, 571 F. Supp. 763, 771 (D. Conn. 1983) (recording that 3.2% of

choice: respondents must indicate that they perceive the asserted mark as a designation of source or not; they must indicate that they are confused as to source or not. The result is that trademark surveys fail to measure potentially valuable empirical information about consumer uncertainty. That is a real shortcoming because social science evidence strongly suggests that the likelihood that a person will take some marketplace action on the basis of a particular belief is tied to the strength with which that belief is held.¹⁷ For example, consumers who strongly believe that a product comes from a company whose products they like are more likely to purchase based on that belief, when compared to a consumer who holds that same belief but only weakly. Thus, the relevance of belief strength to the goals of trademark law should be obvious. And yet had trademark survey experts deliberately set out to conceal consumer uncertainty and dodge the important theoretical questions that uncertainty raises, it is not clear they could have come up with better formats than those currently in use.

In this Article, we argue that trademark law and trademark consumer surveys in particular must finally acknowledge consumer uncertainty—the reality that consumer beliefs are not binary but held at varying levels of strength and meaningfulness. Specifically, the central inquiry in trademark litigation and trademark consumer surveys should ask: Is it likely that some threshold proportion of consumers hold a particular belief *at a substantial level of certainty*? To be sure, incorporating consumer belief strength adds an additional degree of complexity to trademark doctrine and fact-finding, but we think that the benefits in information gained far outweigh the costs.

To defend this claim, we explain why consumer uncertainty matters in trademark law and present experimental evidence of the extent to which conventional survey methods underreport consumer uncertainty. We also describe and test improved survey methods that record consumers' varying levels of confidence in the answers they provide to trademark surveys. We recognize and are sympathetic to the view held by many that even improved consumer surveys would remain, at best, an imperfect form of evidence in trademark disputes. But the fact is that surveys continue to be widely used, and with the shift to faster, less expensive online surveys, their use will likely increase.

respondents to a genericism survey question responded that they did not know or were not sure), *vacated*, 609 F. Supp. 588 (D. Conn. 1985); *see also infra* notes 59–65 and accompanying text (discussing survey respondents' reluctance to admit that they have no opinion or do not know).

¹⁷ *See infra* text accompanying notes 48–54.

This Article pursues two goals. The first goal, as a practical matter, is to detail a low-cost, easily administered, and relatively simple modification to the design of trademark surveys that, by registering consumer uncertainty, will provide courts with substantially better information about consumer beliefs. The second goal is to detail how, as a normative matter, courts should interpret varying levels of consumer uncertainty when adjudicating trademark disputes—not simply to reach the right result in particular cases but more broadly to foster a more efficient and theoretically coherent trademark system and a more competitive marketplace. Trademark law is the youngest of the three main areas of intellectual property law and, in many ways, its doctrine remains in a primitive state. That trademark law still unthinkingly treats consumer beliefs as essentially binary, as all or nothing, is an example of this. We believe it is time that the law advance beyond this simplistic view.

Part I provides background. Our main insight is this: Trademark courts routinely engage in probabilistic reasoning when evaluating whether a *population* of consumers holds a particular belief. However, in doing so, they typically fail to take into account the fact that *individual consumers* themselves engage in probabilistic reasoning about whether various propositions are more or less likely to be true. We briefly review the early history of consumer survey evidence in trademark litigation. We show, remarkably, that when litigants first began a century ago to present survey evidence in trademark disputes, those surveys did exactly what we call upon current surveys to do: they tested not simply for whether a consumer held a particular belief, but also examined how strongly the consumer held that belief. We trace how through the course of the twentieth century trademark litigation surveys declined in both scientific quality and usefulness to courts as they abandoned any commitment to registering the strength of respondents' beliefs.

Part II presents the results of a series of trademark consumer survey experiments we conducted that reveal the extent of uncertainty among survey respondents and demonstrate how to record that uncertainty. We focused on three leading survey formats: the *Teflon* survey format for assessing trademark distinctiveness, and the *Squirt* and *Eveready* survey formats for assessing the likelihood of consumer confusion.¹⁸ Adapting facts from recently litigated cases, we conducted each experiment under two conditions. Under the first condition, we followed current survey methods by providing respondents with a choice

¹⁸ These survey formats are commonly referred to by the name of the cases in which they were first introduced. See *infra* text accompanying notes 111–123 (*Teflon* surveys); 136–151 (*Squirt* surveys); and 163–170 (*Eveready* surveys).

among conventional responses. Under the second condition, we modified the conventional survey formats to allow respondents to indicate their level of confidence in their responses, for example, by means of a seven-point Likert scale. We show that when we restructure surveys to measure uncertainty, we find a lot of it. We also show that providing respondents with answer choices that register their uncertainty produces significant additional information and a richer understanding of consumer belief.¹⁹

Part III sets out the principal implications of our empirical findings for trademark adjudication and policy. First, and most concretely, trademark survey formats must be redesigned to register the strength of respondent belief, such as through the revised trademark survey formats described and demonstrated in Part II. Survey experts should no longer be given license to hide the reality of respondent uncertainty from the finder of fact. Second, courts should take into account the strength of consumer beliefs when determining whether the plaintiff has satisfied its burden of persuasion. Evidence showing, for example, that some proportion of consumers believe it to be only *somewhat likely* that the defendant's mark originates from the plaintiff's should not be the basis for trademark liability, not least because such a weakly held belief may be dispelled when consumers are making decisions in an actual market setting, which almost always provides context that the survey environment lacks. Allowing a plaintiff to use weakly held beliefs to satisfy its burden is especially inappropriate if a larger share of consumers appear to hold the opposite belief with greater certainty. Third, when a court does find liability, it should consider the strength of consumers' mistaken confusion as to the source when tailoring an appropriate remedy. Remedies short of an outright injunction, such as modifications to the defendant's mark or requiring a disclaimer, may be just as effective in disabusing consumers of their weakly held, mistaken beliefs while at the same time limiting the costs imposed on plaintiffs' competitors. Finally, and more broadly, we explain how our empirical findings raise more fundamental normative questions about precisely what kind of harm trademark law is meant to prevent. Should the law police all consumer beliefs, even those held abstractly or notionally? Or should the law concern itself only with consumer beliefs that

¹⁹ The original surveys conducted in their respective litigations had the goal of isolating the impact of the allegedly infringing mark. See *infra* text accompanying notes 111–123 (*Teflon* surveys), 136–151 (*Squirt* surveys), 163–170 (*Eveready* surveys). As such, they employed control treatments in a test-control format in which the presence of the allegedly infringing mark is compared against its absence, all else being equal. See *infra* text accompanying notes 111–123 (*Teflon* surveys), 136–151 (*Squirt* surveys), 163–170 (*Eveready* surveys). Our objective is different. We are interested in isolating the informational impact of assessing consumer uncertainty. Thus, for our objectives, the original survey question without uncertainty assessment serves as a control against which data obtained from questions that assess uncertainty can be compared.

are likely to affect conduct and yield some market consequence? If the latter is correct, as we believe it is, then trademark law requires a more developed account of the conditions in which belief is likely to mature into action, and how the strength with which consumers hold their beliefs will affect their behavior.

I. BACKGROUND: FRAMING CONSUMER UNCERTAINTY IN TRADEMARK LAW

Trademark law boasts a nuanced approach to understanding populations of consumers, but the sophistication of trademark law's empirical grounding is undermined by the law's strangely reductive understanding of the beliefs of individual consumers within those populations. We first discuss here how trademark courts currently analyze consumer perception in the marketplace. We then call upon basic social science learning about consumer perception to explain why trademark law's current approach is inadequate.

A. *Trademark Law and the Probabilism of Consumer Beliefs*

The central purpose of trademark law is to preserve and promote the quality of information in the marketplace—specifically, information about the source and characteristics of products.²⁰ It does so by providing exclusive rights in designations of the source of products. If consumers may rely on these designations of source to accurately identify the products they seek, then their search costs for those products are minimized.²¹ As for producers, those with a reputation for quality can use their trademarks to internalize the rewards from that reputation.²² For this reason, trademark law is highly sensitive to how consumers interpret marketplace information and what beliefs they form about it.²³ Indeed, this Article focuses on consumer beliefs because courts routinely treat those beliefs as dispositive of outcomes in trademark disputes. Consider the most fundamental inquiry in trademark law: should a designation—be it a word, image, shape, sound, or any other perceptible sign—qualify for trademark protection? Courts commonly answer this question by asking whether the relevant consumer population (e.g., car buyers) perceives the designation as an indication of the particular commercial source of the products to which it is

²⁰ See *Lawn Managers, Inc. v. Progressive Lawn Managers, Inc.*, 390 F. Supp. 3d 975, 982 (E.D. Mo. 2018) (“The purpose of federal trademark law is twofold: (1) ‘to protect the public so it may be confident that, in purchasing a product bearing a particular trade-mark which it favorably knows, it will get the product which it asks for and wants to get’ and (2) ‘where the owner of a trade-mark has spent energy, time, and money in presenting to the public the product, he is protected in his investment from its misappropriation[.]’” (alteration in original) (quoting S. Rep. No. 79-1333, at 3 (1946))), *aff’d*, 959 F.3d 903 (8th Cir. 2020).

²¹ See *Qualitex Co. v. Jacobson Prod. Co.*, 514 U.S. 159, 163–64 (1995).

²² *Id.* at 164.

²³ See *Lawn Managers*, 390 F. Supp. 3d at 982.

affixed (e.g., TESLA or PRIUS), and not as a mere description of the product used by many different sources (e.g., “fuel efficient” or “carbon neutral”) or as a generic label for a product category (“EV” or “hybrid car”).²⁴ A similar focus on consumer belief pervades courts’ determinations of the main liability question in trademark law. Even when a designation is deemed a qualifying trademark, courts do not on that basis grant to the mark owner “in gross” property rights against all unauthorized uses.²⁵ Rather, trademark law prohibits only uses that a court determines are likely to cause consumers to be confused as to the true source of the defendant’s products—typically by misleading consumers to believe that the defendant’s products originate from the plaintiff.²⁶ To adjudicate most defenses to infringement in trademark litigation, courts must similarly inquire into the thoughts and beliefs of people who are likely to encounter or to consume the products at issue.²⁷

The degree to which consumer beliefs shape trademark law is perhaps nowhere clearer than in the Supreme Court’s many forays into the area.²⁸ The Court has repeatedly deferred to empirical evidence—or generated its own hypotheses about the empirics of consumer perception—to resolve cases or develop new doctrine. In her opinion for the majority in *Booking.com*,²⁹ Justice Ginsburg relied on the respondent’s survey evidence to override much of the policy discussion in the case about whether the trademark at issue should be deemed generic and thus unprotectable: “Consumers do not in fact perceive the

²⁴ See MCCARTHY, *supra* note 1, § 2:10 (“While a trademark can be categorized as a kind of ‘property’ right, such a characterization often creates more confusion than clarity. This is because the ‘property’ parameters of a trademark are defined very differently from any other kind of ‘property.’ In almost all cases, the exclusive ‘property’ right of a trademark is defined by customer perception.”). In rare instances, courts will rule contrary to consumer perception. For example, even if consumers perceive certain product features as designations of source, courts will refuse to give those features trademark protection if they are functional. *See TrafFix Devices, Inc. v. Mktg. Displays, Inc.*, 532 U.S. 23, 26 (2001) (“[S]econdary meaning is irrelevant because there can be no trade dress protection in any event.”); *see id.* at 33 (“*Functionality* having been established, whether [the plaintiff’s] dual-spring design has acquired secondary meaning need not be considered.”) (emphasis added).

²⁵ *See, e.g.*, *Savannah Coll. of Art & Design, Inc. v. Sportswear, Inc.*, 872 F.3d 1256, 1266 (11th Cir. 2017).

²⁶ MCCARTHY, *supra* note 1, § 2:10 (“Once created, the ‘property’ right of a trademark is invaded by a junior user’s use of a similar mark that is likely to create confusion. Hence, the ‘property’ in a trademark is the right to prevent confusion.”). Note that trademark law also prohibits, under the heading of trademark “dilution,” uses that impair the distinctiveness of a famous mark in the minds of consumers. *See Lanham Act* § 43(c), 15 U.S.C. § 1125(c).

²⁷ For example, a court’s analysis of the descriptive fair use defense typically considers, among other things, whether consumers would perceive the defendant’s conduct as constituting use of the term at issue “other than as a mark” and “in a descriptive sense.” *See, e.g.*, *Dessert Beauty, Inc. v. Fox*, 568 F. Supp. 2d 416, 423 (S.D.N.Y. 2008) (quoting *EMI Catalogue P’ship v. Hill, Holliday, Connors, Cosmopoulos Inc.*, 228 F.3d 56, 64 (2d Cir. 2000)).

²⁸ *See U.S. Pat. & Trademark Off. v. Booking.com B.V.*, 140 S. Ct. 2298 (2020); *Wal-Mart Stores, Inc. v. Samara Bros.*, 529 U.S. 205 (2000); *Dastar Corp. v. Twentieth Century Fox Film Corp.*, 539 U.S. 23 (2003).

²⁹ *Booking.com*, 140 S. Ct. at 2300–01.

term ‘Booking.com’ [as generic] That should resolve this case: Because ‘Booking.com’ is not a generic name to consumers, it is not generic.”³⁰ For his part, Justice Scalia was criticized (or admired) for resolving various knotty trademark issues before the Court through sweeping empirical claims in his opinions about consumer perception.³¹

Given the decisive role that consumer perceptions play in the outcome of trademark disputes, it is of the utmost importance that courts understand what consumers actually believe. To do so, courts typically address their analysis to *populations* of relevant consumers and assess those populations probabilistically.³² The “likelihood of confusion” cause of action prompts courts to ask whether the defendant’s trademark is likely to cause the relevant population of consumers to mistakenly believe that the plaintiff’s and defendant’s goods originate from the same source.³³ With respect to the “likelihood of dilution” cause of action, courts ask whether the defendant’s trademark is likely to diminish the distinctiveness of the plaintiff’s famous mark in the minds of a relevant population of consumers.³⁴ The antecedent question of distinctiveness—i.e., whether an asserted mark is protectable at all—is implicitly framed in the same way: to establish the distinctiveness of a

³⁰ *Id.* at 2305.

³¹ For example, in *Wal-Mart Stores*, 529 U.S. 205, the Court based its holding on its own hypothesis about how consumers think. Writing for the Court, Justice Scalia explained:

In the case of product design, . . . we think consumer predisposition to equate the feature with the source does not exist. Consumers are aware of the reality that, almost invariably, even the most unusual of product designs—such as a cocktail shaker shaped like a penguin—is intended not to identify the source, but to render the product itself more useful or more appealing.

Id. at 213. Thus, the Court held that while a product’s design could qualify for trademark protection, it could only do so if the producer could show that consumers had over time come to believe that the producer was the only source of the particular design—i.e., that the design had developed “acquired distinctiveness” in the minds of consumers as a designation of source. *Id.* at 211, 215–16; *see also Dastar*, 539 U.S. at 32–33 (“The words of the Lanham Act should not be stretched to cover matters that are typically of no consequence to purchasers.”).

³² See Barton Beebe & C. Scott Hemphill, *The Scope of Strong Marks: Should Trademark Law Protect the Strong More than the Weak?*, 92 N.Y.U. L. REV. 1340, 1347–49 (2017).

³³ For federally registered marks, Section 32 of the Lanham Act brands a defendant’s use as actionable trademark confusion if it is “likely to cause confusion, or to cause mistake, or to deceive.” Lanham Act § 32, 15 U.S.C. § 1114(1). For unregistered marks, Section 43(a) of the Lanham Act defines an infringing use as one “likely to cause confusion, or to cause mistake, or to deceive as to the affiliation, connection, or association” of the junior user with the senior user. Lanham Act § 43(a), 15 U.S.C. § 1125(a).

³⁴ See Lanham Act § 43(c), 15 U.S.C. § 1125(c).

descriptive term³⁵ or an element of product design trade dress,³⁶ courts consider how likely it is that a substantial proportion of the relevant consumer population perceives the term or element as distinctive of source.

Importantly, in assessing consumer beliefs, trademark law recognizes that most populations of relevant consumers are not homogenous.³⁷ That is why the likelihood of confusion and likelihood of dilution causes of action do not require courts to find that it is likely that the *entire* population of relevant consumers is confused or experiencing dilution.³⁸ Instead, trademark law asks courts to look inside the population of relevant consumers and determine whether an appreciable proportion of that population (typically, 20–25%,³⁹ but sometimes as low as 15%⁴⁰) is confused or experiencing dilution. If there is a better-than-even chance that the defendant’s conduct will confuse or dilute more than that threshold proportion, then a court should find infringement. Trademark law takes the same approach when inquiring whether a particular designation functions as a mark in the first place—though the threshold is typically set higher.⁴¹ As with likelihood of confusion, tests for trademark genericism and

³⁵ Descriptive terms are protectable as marks if the plaintiff establishes that they have acquired distinctiveness (sometimes referred to as “secondary meaning”); i.e., that an appreciable number of consumers perceive them as indicating the source of particular products or services. *Qualitex Co. v. Jacobson Prod. Co.*, 514 U.S. 159, 163 (1995). A class of “inherently distinctive” marks—i.e., fanciful, arbitrary, and suggestive marks—are protected without the need for plaintiff to establish distinctiveness. *See Abercrombie & Fitch Co. v. Hunting World, Inc.*, 537 F.2d 4, 9–10 (2d Cir. 1976).

³⁶ *See Wal-Mart Stores, Inc.*, 529 U.S. at 216 (“[I]n an action for infringement of unregistered trade dress under § 43(a) of the Lanham Act, a product’s design is distinctive, and therefore protectible, only upon a showing of secondary meaning.”).

³⁷ MCCARTHY, *supra* note 1, § 32:185 (noting how courts look for an “appreciable” number of consumers, which may not even be a majority).

³⁸ *Id.*

³⁹ *See, e.g.*, McDonald’s Corp. v. McBagel’s, Inc., 649 F. Supp. 1268, 1277 (S.D.N.Y. 1986) (showing that 25% supports finding of likely confusion); Bell v. Starbucks U.S. Brands Corp., 389 F. Supp. 2d 766, 776 (S.D. Tex. 2005) (finding 25% to be sufficient to show a “significant” level of actual confusion and to support a finding of infringement), *aff’d*, 205 F. App’x 289 (5th Cir. 2006); *see also* MCCARTHY, *supra* note 1, § 32:188 (“Generally, figures in the range of 25% to 50% have been viewed as solid support for a finding of a likelihood of confusion. . . . In the author’s view, survey confusion numbers that go below 20% need to be carefully viewed against the background of other evidence weighing for and against a conclusion of likely confusion.”).

⁴⁰ *Exxon Corp. v. Tex. Motor Exch., Inc.*, 628 F.2d 500, 507 (5th Cir. 1980) (explaining that a survey showing 15% confusion was “strong evidence” of a likelihood of confusion where other evidence was also strongly supportive); *see also* MCCARTHY, *supra* note 1, § 32:188 (reviewing case law relying on a 15% rate of confusion in survey evidence as probative of likely confusion).

⁴¹ *See, e.g.*, Spraying Sys. Co. v. Delavan, 975 F.2d 387, 394 (7th Cir. 1992) (“While a 50-percent figure is regarded as clearly sufficient to establish secondary meaning, a figure in the thirties can only be considered marginal.”).

distinctiveness base their findings on the percentage of the relevant consumer population that perceives the indicium at issue as indicating source.⁴²

That said, although trademark law recognizes the heterogeneity of beliefs within a given population of consumers, the empirical sophistication of trademark law stops there. It does not go deeper to consider the strength and meaningfulness of the beliefs held by each individual within that population. For example, while trademark law's likelihood of confusion analysis assesses consumer populations in continuous terms as proportionally more or less confused, it typically assesses individuals within those populations as binaries; each is either absolutely confused or absolutely not confused. The same is true for trademark law's distinctiveness analysis. It assesses consumer populations in continuous terms as manifesting a proportionally higher or lower incidence of belief that a particular asserted mark indicates the source of a product. But it treats individuals within those populations as binaries: each individual either totally supports or totally rejects the proposition that the asserted mark indicates that source.⁴³

At the foundation of trademark law is thus an unrealistic simplification of the individual beliefs that, in the aggregate, determine protectability and liability in trademark cases. Trademark law typically treats an individual consumer's subjective probability as either 1 or 0, then adds up the 1s in a relevant population, and from that derives a proportion of consumers who either do or do not hold a particular belief. In contrast, the social science literature has long recognized the obvious: individual beliefs are not binaries.

B. The Social Science of Consumer Beliefs

The leading trademark survey formats fail to incorporate important insights from social science research. In this section we examine two of those insights: (1) belief strength affects human behavior; and (2) some survey respondents have a tendency to express beliefs that they do not meaningfully hold. This section develops the first insight in the context of the expectancy value or

⁴² See Wal-Mart Stores, Inc. v. Samara Bros., Inc., 529 U.S. 205, 211 (2000) ("[A] mark has acquired distinctiveness, even if it is not inherently distinctive, if it has developed secondary meaning, which occurs when, 'in the minds of the public, the primary significance of a [mark] is to identify the source of the product rather than the product itself.'" (quoting Inwood Labs., Inc. v. Ives Labs., Inc., 456 U.S. 844, 851, n.11 (1982) (alteration in original))).

⁴³ See Itamar Simonson, *Trademark Infringement from the Buyer Perspective: Conceptual Analysis and Measurement Implications*, 13 J. PUB. POL. & MKTG. 181, 195 (1994) (noting that trademark surveys typically fail to account for respondents' degree of confidence in their responses).

Fishbein model. It then explains the second, which this Article refers to as the “nonattitudes” problem, and discusses how it manifests in trademark surveys.

The concept of belief strength is fundamental to social science understandings of consumer perception and behavior. Psychologists have defined beliefs as linkages between objects and their attributes.⁴⁴ In the context of trademark law, the “object” is the perceptible form of the trademark and the main “attribute” of interest is its reference to a particular commercial source. Just as individuals generally will have varying judgments of “the perceived likelihood that [an] object has (or is associated with) the attribute in question,”⁴⁵ individual consumers will also be more or less certain that an asserted trademark refers to a particular source. Behavioral scientists recommend that “‘belief strength,’ or more simply ‘belief,’ be measured by a procedure which places the subject along a dimension of subjective probability involving [the] object and some related attribute.”⁴⁶ The effect of beliefs on behavior is complex and moderated by many factors, including personality traits, situational factors, and social influences.⁴⁷ Numerous studies indicate, however, that, in general, beliefs become more salient predictors of behavior the better developed and more strongly held they are.⁴⁸

The extensive social psychology literature on consumers’ “attitude construct” highlights the importance of belief strength to consumer behavior.⁴⁹ This literature is based on the assumption that consumers’ attitudes toward competing brands are important determinants of consumer purchasing

⁴⁴ MARTIN FISHBEIN & ICEK AJZEN, BELIEF, ATTITUDE, INTENTION, AND BEHAVIOR: AN INTRODUCTION TO THEORY AND RESEARCH 12 (1975).

⁴⁵ *Id.*

⁴⁶ *Id.*

⁴⁷ See generally Stephen J. Kraus, *Attitudes and the Prediction of Behavior: A Meta-Analysis of the Empirical Literature*, 21 PERSONALITY & SOC. PSYCHOL. BULL. 58 (1995).

⁴⁸ See Jon A. Krosnick & Robert P. Abelson, *The Case for Measuring Attitude Strength in Surveys*, in QUESTIONS ABOUT QUESTIONS: INQUIRIES INTO THE COGNITIVE BASES OF SURVEYS 177–78 (Judith M. Tanur, ed., Russell Sage Found., 1992); Denis T. Regan & Russell Fazio, *On the Consistency Between Attitudes and Behavior: Look to the Method of Attitude Formation*, 13 J. EXPERIMENTAL SOC. PSYCH. 28, 30 (1977); Charles R. Tittle & Richard J. Hill, *Attitude Measurement and Prediction of Behavior: An Evaluation of Conditions and Measurement Techniques*, 30 SOCIOOMETRY 199, 199 (1967); see also Russell H. Fazio & Mark P. Zanna, *On the Predictive Validity of Attitudes: The Roles of Direct Experience and Confidence*, 46 J. PERSONALITY 228, 228–42 (1978) (reporting the results of an experiment involving puzzle games showing that participants’ indications on an eleven-point scale of high levels of confidence in their initial judgments of interest in a game were highly predictive of participants’ subsequent behavior with respect to the game); John Sample & Rex Warland, *Attitude and Prediction of Behavior*, 51 SOC. FORCES 292 (1973) (showing that students’ indications of high levels of certainty in their responses to survey about interest in student government were robust predictors of students’ voting behavior).

⁴⁹ Icek Ajzen, *Consumer Attitudes and Behavior*, in HANDBOOK OF CONSUMER PSYCHOLOGY 525, 530 (Curtis P. Haugvedt et al. eds., 2008).

decisions.⁵⁰ Attitudes are generally defined as the tendency to respond to an object (e.g., a brand) either more or less favorably or unfavorably.⁵¹ More favorable attitudes lead to more positive intentions. Indeed, social psychologists view intentions as mediating the relationship between attitudes and behavior (e.g., purchasing).⁵² Consumer researchers have conceptualized this causal sequence as the belief-attitude-intention hierarchy.⁵³ Beyond that, others have established that consumer intentions drive actual consumer purchasing decisions.⁵⁴

One leading social psychology model of consumer decision-making further emphasizes just how much importance social science places on belief strength in understanding consumer perception and consumer behavior. Belief strength is defined in the context of consumer behavior as the subjective probability that a product or brand possesses a certain attribute (e.g., that it comes from a specific source).⁵⁵ The expectancy-value model, such as that proposed by Martin

⁵⁰ See, e.g., Frédéric F. Brunel, Brian C. Tietje & Anthony G. Greenwald, *Is the Implicit Association Test a Valid and Valuable Measure of Implicit Consumer Social Cognition?*, 14 J. CONSUMER PSYCH. 385, 385–404 (2004); Keith Coulter & Girish Punj, *The Effects of Cognitive Resource Requirements, Availability, and Argument Quality on Brand Attitudes: A Melding of Elaboration Likelihood and Cognitive Resource Matching Theories*, 33 J. ADVERT. 53, 53–64 (2004); Jaideep Sengupta & Gavan J. Fitzsimons, *The Effect of Analyzing Reasons on the Stability of Brand Attitudes: A Reconciliation of Opposing Predictions*, 31 J. CONSUMER RSCH. 705, 705–11 (2004).

⁵¹ See ALICE EAGLY & SHELLY CHAIKEN, *THE PSYCHOLOGY OF ATTITUDES* 1 (1993); FISHBEIN & AJZEN, *supra* note 44, at 10; CHARLES E. OSGOOD, *THE MEASUREMENT OF MEANING* 189–91 (1957). See generally Richard E. Petty & John T. Cacioppo, *The Elaboration Likelihood Model of Persuasion*, 19 ADVANCES IN EXPERIMENTAL SOC. PSYCH. 123 (1986).

⁵² See Richard P. Bagozzi & Paul R. Warshaw, *Trying to Consume*, 17 J. CONSUMER RSCH. 127, 127–40 (1990); Peter M. Bentler & George Speckart, *Models of Attitude-Behavior Relations*, 86 PYSCH. REV. 452, 452–64 (1979); FISHBEIN & AJZEN, *supra* note 44, at 12–13; Jeffrey D. Fisher & William A. Fisher, *Changing AIDS-Risk Behavior*, 111 PYSCH. BULL. 455, 455–74 (1992); Peter M. Gollwitzer, *Goal Achievement: The Role of Intentions*, 4 EUR. REV. SOC. PSYCH. 141, 141–85 (1993); Julius Kuhl, *Volitional Aspects of Achievement Motivation and Learned Helplessness: Toward a Comprehensive Theory of Action Control*, 13 PROGRESS IN EXPERIMENTAL PERSONALITY RSCH. 99, 99–171 (1985); EDWIN A. LOCKE & GARY P. LATHAM, *A THEORY OF GOAL SETTING AND TASK PERFORMANCE* 6–9, 13 (1990); HARRY C. TRIANDIS, *INTERPERSONAL BEHAVIOR* 9 (1977).

⁵³ See Scott B. Follows & David Jobber, *Environmentally Responsible Purchase Behavior: A Test of a Consumer Model*, 34 EUR. J. MKTG. 723, 723–46 (2000); Robert Madrigal, *Social Identity Effects in a Belief-Attitude Intentions Hierarchy: Implications for Corporate Sponsorship*, 18 PSYCH. & MKTG. 145, 145–65 (2001); Jennifer Paff Ogle, Karen H. Hyllegard & Brian H. Dunbar, *Predicting Patronage Behaviors in a Sustainable Retail Environment: Adding Retail Characteristics and Consumer Lifestyle Orientation to the Belief-Attitude-Behavior Intention Model*, 36 ENV'T & BEHAV. 717, 717–47 (2004).

⁵⁴ See, e.g., Vicki Morwitz, Joel Steckel & Alok Gupta, *When Do Purchase Intentions Predict Sales?*, 23 INT'L J. FORECASTING 347, 347–64 (2007).

⁵⁵ FISHBEIN & AJZEN, *supra* note 44, at 12.

Fishbein,⁵⁶ holds that people's attitudes toward a product are based on their evaluation of the potential attributes of that product weighted by the subjective probability that the product actually possesses that attribute. In other words, a consumer's attitude toward a product is a weighted sum of the consumer's evaluation of the product's attributes where the weights are beliefs about those attributes.⁵⁷ Mathematically we can write:

$$A \propto \sum_{i=1}^n b_i e_i$$

where A is the consumer's attitude toward the product; b_i is the consumer's belief that the product has attribute i ; e_i is the consumer's evaluation of attribute i ; and n is the number of product attributes. Marketing scholars have thoroughly established the validity of this model.⁵⁸

Remarkably, trademark law and trademark surveys in particular currently ignore even these basic (and noncontroversial) social science findings. In Fishbein's model, the variable b_i is the continuous variable denoting the consumer's subjective probability that the product originates from the source i . If trademark law thinks of this construct at all, it thinks of it as binary. The result is that trademark law and trademark surveys typically fail to distinguish between consumers who hold meaningful beliefs likely to translate into some appreciable effect on the owner of a mark and those whose ambivalence will not lead to action or market harm.

Finally, social science research points to an additional significant problem with the failure of current trademark survey protocols to test for consumer belief strength: the problem of "nonattitudes."⁵⁹ Trademark surveys are cognitively

⁵⁶ Martin Fishbein, *An Investigation of the Relationships Between Beliefs About an Object and the Attitude Toward that Object*, 16 HUM. RELS. 233, 233–40 (1963); Martin Fishbein, *A Consideration of Beliefs and Their Role in Attitude Measurement*, in *READINGS IN ATTITUDE THEORY AND MEASUREMENT* 257–66 (1967).

⁵⁷ See *supra* note 56.

⁵⁸ See Richard J. Lutz, *An Experimental Investigation of Causal Relations Among Cognitions, Affect, and Behavioral Intention*, 3 J. CONSUMER RSCH. 197, 197–208 (1977); Andrew A. Mitchell & Jerry C. Olson, *Are Product Attribute Beliefs the Only Mediator of Advertising Effects on Brand Attitude?*, 18 J. MKTG. RES. 318, 318–32 (1977).

⁵⁹ See Philip E. Converse, *Attitudes and Non-Attitudes: Continuation of a Dialogue*, in *THE QUANTITATIVE ANALYSIS OF SOCIAL PROBLEMS* 168–89 (Edward Tufte ed. 1970); see also Lee Sigelman & Dan Thomas, *Opinion Leadership & the Crystallization of Nonattitudes: Some Experimental Results*, 16 POLITY 484, 484 (1984) ("What is it that prompts as many as one respondent in three to express an opinion, pro or con, on a given issue with absolutely no information or knowledge on the matter to guide his or her response?").

demanding tasks that require respondents to quickly generate beliefs about the origins of products that the respondents may never have seen or previously thought about.⁶⁰ Ideally, survey respondents who are unable to form meaningful opinions will say so by selecting options such as “don’t know.” But social science studies have long made clear that significant numbers of survey respondents are unwilling to admit that they do not know or have no opinion.⁶¹ Instead, they engage in what Jon Krosnick has described as “mental coin-flipping” and select answer choices at random.⁶² These nonattitudes or “pseudo-opinions”⁶³ look like valid responses, but they do not measure an underlying attitude or meaningful belief.⁶⁴

There is a substantial danger, in other words, that trademark surveys may prompt some respondents to provide responses unrelated to their actual marketplace beliefs, if they have any, on a particular question. Other respondents may hold multiple conflicting beliefs but are nevertheless forced to express just one. Still others, though they may not be guessing or choosing at random, may be uncertain and produce significant “response instability”⁶⁵—if asked the same question at a later time, they may respond differently.

The social science learning on belief strength and nonattitudes reviewed here prompts a number of straightforward research questions relating to trademark surveys that can be answered experimentally. Do trademark survey respondents have varying levels of confidence in their survey responses? Do respondents offer responses that take the form of nonattitudes? Are there workable methods

⁶⁰ See JACOB JACOBY, TRADEMARK SURVEYS VOLUME I: DESIGNING, IMPLEMENTING, AND EVALUATING SURVEYS 151–53 (2013).

⁶¹ See Philip E. Converse, *The Nature of Belief Systems in Mass Publics*, 18 CRITICAL REV. 1, 49 (1964); see also John P. Liefeld, *How Surveys Overestimate the Likelihood of Consumer Confusion*, 93 TRADEMARK REP. 939 (2003) (in light of survey respondents’ reluctance to state that they have no opinion or do not know, reporting the results of a series of experiments involving trademark surveys using different forms of “filter questions” asking if respondents had a previously formed opinion or attitude available in memory to elicit no opinion or do not know responses).

⁶² Jon A. Krosnick, *Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys*, 5 APPLIED COGNITIVE PSYCH. 213, 220 (1991).

⁶³ See Sigelman & Thomas, *supra* note 59, at 484.

⁶⁴ English courts are especially sensitive to the problem of nonattitudes in trademark survey evidence and explicitly require surveys to avoid prompting respondents to form beliefs that they would not otherwise have had. According to the “Whitford Guidelines” developed by Mr. Justice Whitford in *Imperial Group plc & Another v. Philip Morris Ltd. & Another*, [1984] RPC 293, a survey question must not “direct the person answering the question into a field of speculation upon which that person would never have embarked had the question not been put.” *Id.* at 303; see also *Interflora Inc. v. Marks & Spencer Plc*, [2013] FSR 21, [151] (requiring courts to consider “[e]vidence that any further survey will comply with the Whitford guidelines” when determining whether to grant a party permission to conduct a survey).

⁶⁵ John Zaller & Stanley Feldman, *A Simple Theory of the Survey Response: Answering Questions Versus Revealing Preferences*, 36 AM. J. POL. SCI. 579, 580 (1992).

by which trademark surveys can test for belief strength and nonattitudes? Part II presents empirical evidence that addresses each of these questions. But first, Section C briefly reviews the origins and early development of trademark surveys because these questions once figured prominently in the design and analysis of this form of evidence.

C. The Treatment of Consumer Uncertainty in the Early History of Trademark Survey Evidence

It is unfortunate that the early history of trademark survey evidence has been largely forgotten, because there is much we can learn—or relearn—from it. The trademark survey formats first proposed by social scientists a century ago were, in general, scientifically superior to those in use now. This is because they tested for consumer belief strength, b_i in the expectancy-value model explained above.⁶⁶ The story of trademark survey evidence over the past century is a story of regression to the blunt instruments used today.

The first survey evidence ever submitted to an American court in a trademark dispute was introduced by Coca-Cola in 1921.⁶⁷ In 1915, Coca-Cola opposed the registration of the mark CHERO-COLA for cola-flavored soft drinks, asserting that it was confusingly similar with the mark COCA-COLA.⁶⁸ Coca-Cola submitted a survey report as evidence that consumers would confuse the marks.⁶⁹ In the report, the Columbia-University-affiliated psychologist Richard Paynter described four experiments he conducted under laboratory conditions.⁷⁰ In the first of these experiments, the respondent was shown twenty slips of paper, each of which displayed a typed word mark.⁷¹ One of the slips of paper bore the mark “Coca-Cola.”⁷² After a brief pause, the respondent was then shown in random order forty slips of paper, twenty of which bore marks not previously

⁶⁶ See *supra* notes 56–62 and accompanying text.

⁶⁷ See *Coca-Cola Co. v. Chero-Cola Co.*, 273 F. 755, 756 (D.C. Cir. 1921).

⁶⁸ See Edward S. Rogers, *An Account of Some Psychological Experiments on the Subject of Trade-Mark Infringement*, 18 MICH. L. REV. 75, 77 (1919).

⁶⁹ See *id.*

⁷⁰ See *id.* at 77–99 (reproducing the report); see also Richard H. Paynter, *A Psychological Study of Trade-Mark Infringement*, 42 ARCHIVES PSYCH. 1, 6 (1920) (discussing experiments related to trademark infringement); Richard H. Paynter, *A Psychological Study of Confusion Between Word Trade-Marks*, 11 BULL. U.S. TRADE-MARK ASSOC. 101, 101 (1915) (reporting the results of experiments similar to those Paynter used in *Coca-Cola Co. v. Chero-Cola Co.*) (Paynter’s experiments were similar to those envisioned, but not carried out, by Hugo Münsterberg. See HUGO MÜNSTERBERG, *PSYCHOLOGY AND INDUSTRIAL EFFICIENCY* 282–93 (1913)).

⁷¹ See Rogers, *supra* note 68, at 78. The slips were presented in random order except that “Coca-Cola” and “Chero-Cola” appeared neither first nor last. See *id.* at 80.

⁷² *Id.* at 78–79.

presented to the respondent, nineteen of which bore marks previously shown to the respondent, and one of which bore the mark “Chero-Cola” instead of “Coca-Cola.”⁷³ For these forty slips of paper, the written instructions provided to each respondent explained:

[Y]ou will be . . . asked to pick out those marks you have just seen in the presentation and those which you have not seen. You will be further asked to sort the marks into six piles, according to the degree of your confidence or certainty of your recognition of your marks. There are three degrees of certainty for the marks that are recognized as seen, and three similar degrees for those that are recognized as not seen. The three degrees are “absolutely certain,” “reasonably certain,” and a “faint idea.”⁷⁴

Paynter reported that 28% of respondents in the first experiment mistook “Chero-Cola” for “Coca-Cola.”⁷⁵ Paynter further detailed the degree of certainty each respondent indicated in their answer, with 81.8% of those confused in the first experiment stating that they were “absolutely certain.”⁷⁶ In its 1921 opinion in *Coca-Cola Co. v. Chero-Cola Co.*, the United States Court of Appeals for the District of Columbia affirmed the Patent Office’s refusal to register the CHERO-COLA mark.⁷⁷ The court noted that “we are invited to listen to the teaching of psychology on the subject” of consumer confusion but did not explicitly rely on Paynter’s survey.⁷⁸

Soon after the *Chero-Cola* case, a trademark litigant once again submitted survey evidence in which respondents were asked to specify their degree of certainty.⁷⁹ This time, the survey expert was Harold Burtt, an Ohio State-University-affiliated psychologist.⁸⁰ Burtt roughly followed Paynter’s protocols

⁷³ *Id.* at 79. The marks were capitalized but not typed in all upper-case characters. *Id.*

⁷⁴ *Id.* at 80. Paynter’s second and third experiments followed a similar protocol, except that the second experiment included on each slip of paper below the mark the product category for which the mark was used (e.g., “Soft Drink”) and the third experiment included, for purposes of comparison, marks and their products from various recently litigated trademark cases. *Id.* at 79, 83. The fourth experiment exposed the respondents to ten pairs of marks that were the subject of recent trademark infringement cases and asked the respondents to order the pairs according to the degree of confusion that the respondent believed each pair would create in consumers. *Id.* at 93.

⁷⁵ *Id.* at 82. Moreover, Paynter reported that 68% of respondents in the second experiment, in which both slips of paper carried the additional information “Soft Drink,” mistook “Chero-Cola” for “Coca-Cola.” *Id.*

⁷⁶ *Id.* In the second experiment, 70.5% of those confused stated they were “absolutely certain.” *Id.* at 83.

⁷⁷ 273 F. 755, 757 (D.C. Cir. 1921).

⁷⁸ *Id.* at 756.

⁷⁹ See generally Harold E. Burtt, *Measurement of Confusion Between Similar Trade Names*, 19 ILL. L. REV. 320 (1924).

⁸⁰ See *id.* at 320 n.1.

but used a seven-point scale of certainty.⁸¹ The survey showed confusion between the marks at issue and a high degree of certainty including by respondents who were confused.⁸² Yet in its 1923 opinion in *Citizens' Wholesale Supply Co. v. Downing*,⁸³ the Supreme Court of Ohio affirmed the lower court's finding of no infringement without commenting on the survey evidence.⁸⁴

In the decades following *Chero-Cola* and *Citizens' Wholesale*, trademark survey methods shifted primarily to face-to-face interviews with consumers conducted either door-to-door or by intercepting consumers in or outside stores.⁸⁵ None of these interview-based surveys appear to have probed respondents for their degree of certainty in their response, perhaps because in most cases the trademark owner was the party who submitted the survey and would not likely have benefitted from data showing respondent uncertainty.⁸⁶ That said, from time to time courts picked up on the problem. Indeed, in one of the most influential judicial analyses of trademark survey evidence in the midcentury, the court in *General Motors Corp. v. Cadillac Marine & Boat Co.* criticized the plaintiff's survey for failing to "take into consideration the hazy and qualified answers" of the survey's respondents.⁸⁷ General Motors produced automobiles under the mark CADILLAC; Cadillac Marine & Boat produced boats under the same mark.⁸⁸ In an early version of what became the *Eveready* survey format, General Motors asked, "Who do you think puts out the boat shown on the opposite page?" and, "Will you please name anything else put out by the same concern?"⁸⁹ The court closely scrutinized the survey respondents' answers and repeatedly noted respondents' "unclear or ambiguous" responses.⁹⁰

⁸¹ *Id.* at 325.

⁸² *Id.* at 327.

⁸³ 140 N.E. 683 (Ohio 1923).

⁸⁴ *See id.* at 683.

⁸⁵ *See, e.g.*, Lerner Stores Corp. v. Lerner, 162 F.2d 160, 162 (9th Cir. 1947) (discussing an intercept survey conducted in front of the plaintiff's store); DuPont Cellophane Co. v. Waxed Prods. Co., 6 F. Supp. 859, 877 (E.D.N.Y. 1934) (discussing a house-to-house survey testing whether respondents perceived cellophane as a generic term); Oneida, Ltd. v. Nat'l Silver Co., 25 N.Y.S.2d 271, 286 (Sup. Ct. 1940) (discussing two house-to-house surveys); *see also* Beverly W. Pattishall, *Reaction Test Evidence in Trade Identity Cases*, 49 TRADEMARK REP. 145, 156 (1959) (arguing that fixed form interviews are the best means of testing for likelihood of consumer confusion); Robert C. Sorensen & Theodore C. Sorensen, *The Admissibility and Use of Opinion Research Evidence*, 28 N.Y.U. L. REV. 1213, 1215–16 (1953) (arguing that personal interviews are most useful technique of determining public opinion); Robert Bonyng, *Trademark Surveys and Techniques and Their Use in Litigation*, 48 ABA J. 329, 329–30 (1962) (reviewing mid-twentieth century trademark survey methods).

⁸⁶ *See, e.g.*, Int'l Milling Co. v. Robin Hood Popcorn Co., Inc., 110 U.S.P.Q. 368, at *1–2 (Comm'r Pat. & Trademarks 1956).

⁸⁷ 226 F. Supp. 716, 736 (W.D. Mich 1964).

⁸⁸ *Id.* at 719–20.

⁸⁹ *Id.* at 734 n.16.

⁹⁰ *Id.* at 735.

One of the examples the court gave was the response “Well, since it says ‘Cadillac,’ I guess it’s Cadillac.”⁹¹ The survey’s questions had not probed for uncertainty, but the respondents’ verbatim responses revealed it anyway.⁹²

Ultimately, the court in *Cadillac Marine* rejected the plaintiff’s efforts to do what we believe so many current surveys seek to do: hide respondent uncertainty behind bottom-line, summary percentages of those confused and not confused.⁹³ “Such qualified answers,” the court explained, “are not susceptible to a categorization such as plaintiff attempted in summarizing the poll.”⁹⁴ Other courts of the time were similarly critical of trademark survey evidence,⁹⁵ and even as late as the early 1970s, courts remained generally hostile to it.⁹⁶

Things changed dramatically with the Seventh Circuit’s 1976 opinion in *Union Carbide Corp. v. Ever-Ready Inc.*, which largely inaugurated the current era in which survey evidence plays a much more substantial role in trademark litigation.⁹⁷ Union Carbide was the longtime manufacturer of batteries under the mark EVERREADY.⁹⁸ The defendant began importing lamp bulbs and lamps bearing the mark EVER-READY.⁹⁹ Union Carbide produced two surveys, one focusing on light bulbs, the other on lamps.¹⁰⁰ Each survey followed what we

⁹¹ *Id.*

⁹² See *id.* at 735–36.

⁹³ See *id.* at 738.

⁹⁴ *Id.* at 735. The *Cadillac Marine* court further criticized the second main question as leading. In the court’s view, it prompted respondents who “drew a complete blank” on the first question eventually to think of General Motors: “One individual said, ‘I have no idea,’ in answer to the first question yet the second question brought the answer, ‘car.’” *Id.* at 736.

⁹⁵ See, e.g., *Nat’l Biscuit Co. v. Princeton Mining Co., Inc.*, 137 U.S.P.Q. 250, at *3 (T.T.A.B. 1963). In rejecting the plaintiff’s survey, the court noted that:

[A] review of the survey sheets from which the summary was prepared discloses that the figure in question includes many persons who named opposer or its products only after prefacing their answers with such statements as “I have no idea,” “I haven’t the slightest idea,” “You’ve got me,” “Well golly, I don’t know.” and the like

Id.

⁹⁶ See, e.g., *Am. Basketball Ass’n v. AMF Voit, Inc.*, 358 F. Supp. 981, 986 (S.D.N.Y. 1973) (assessing the plaintiff’s secondary meaning survey as unworthy of “any substantial weight”); *Sears, Roebuck & Co. v. Allstate Driving Sch., Inc.*, 301 F. Supp. 4, 18 (E.D.N.Y. 1969) (“One of the dangers inherent in a consumer reaction test is that it is not administered in the context of the market place. Respondents to such a test do not consider those factors which are relevant to the particular purchasing decision at hand.”); *Aerojet-Gen. Corp. v. Cincinnati Screen Process Supplies, Inc.*, No. 6315, 1971 WL 16663, at *6 (S.D. Ohio Apr. 15, 1971) (assessing the defendant’s likelihood of confusion survey as entitled to “very little weight”).

⁹⁷ 531 F.2d 366 (7th Cir. 1976).

⁹⁸ *Id.* at 370.

⁹⁹ *Id.* at 371.

¹⁰⁰ *Id.* at 385.

now identify as the *Eveready* survey format, which is discussed further below.¹⁰¹ Following *Cadillac Marine*, the district court focused on the uncertainty of the survey respondents' answers and the leading nature of the surveys' questions.¹⁰² It concluded that the surveys were entitled to "little, if any, weight" and ultimately found no likelihood of confusion.¹⁰³ The Seventh Circuit reversed.¹⁰⁴ It placed great weight on the survey evidence, going so far as to explain that it would not have reversed on the likelihood of confusion issue "were it not for the survey evidence presented at trial."¹⁰⁵ The court distinguished the survey in *Cadillac Marine* as "sloppy."¹⁰⁶ Rejecting the district court's concerns about respondent uncertainty, the Seventh Circuit held "the district court clearly erroneous in not crediting the surveys taken by Carbide."¹⁰⁷

The influence of the Seventh Circuit's *Ever-Ready* decision was profound. Since *Ever-Ready*, the issue of the degree of respondent certainty, so important in the earliest days of trademark survey evidence, appears to have been more or less forgotten.

II. EXPERIMENTS ASSESSING THE EFFECT OF CONSUMER UNCERTAINTY ON TRADEMARK SURVEY METHODS

To study the extent to which current trademark survey formats obscure respondent uncertainty and nonattitudes, we undertook a series of experiments that built upon surveys submitted in recent trademark litigation. We began by replicating the structure of these surveys based on information provided by the experts who designed them.¹⁰⁸ We then introduced changes to the question format and the answer choices. We also adapted facts from a recent trademark

¹⁰¹ *Id.*

¹⁰² Union Carbide Corp. v. Ever-Ready Inc., 392 F. Supp. 280, 291–94 (N.D. Ill. 1975), *rev'd*, 531 F.2d 366.

¹⁰³ *Id.* at 292, 294.

¹⁰⁴ *Ever-Ready, Inc.*, 531 F.2d at 389.

¹⁰⁵ *Id.* at 385.

¹⁰⁶ *Id.* at 386.

¹⁰⁷ *Id.* at 387. Regrettably, the appellate court also adopted the convention of citing survey percentages from previous cases in which confusion was found and noting that the plaintiff's confusion percentages were higher. *Id.* at 386. Doing so stripped out all detail and nuance from the previous courts' factual findings and reduced the previous cases to simple—and often misleading—summary statistics.

¹⁰⁸ Surveys such as these are generally conducted in a test-control format, which uses a control group to control for alternative explanations for the test group's responses, such as baseline levels of confusion, guessing, and noise. *See infra* text accompanying notes 111–123 (*Teflon* surveys), 136–151 (*Squirt* surveys), 163–170 (*Eveready* surveys). Because we are comparing testing for uncertainty to not testing for uncertainty, we do not need to employ the test-control format of conventional trademark consumer surveys.

dispute¹⁰⁹ to perform additional testing. In total, we conducted experiments for each of the three leading trademark survey formats: the *Teflon* survey format for genericism, the *Squirt* survey format for likelihood of consumer confusion as to source, and the *Eveready* survey format also for likelihood of confusion.¹¹⁰ Our experiments show that current survey formats using binary answer choices fail to detect the existence of high levels of uncertainty and nonattitudes in consumers' responses. Our experiments also show that measuring and calculating uncertainty requires neither complex adaptations nor a complete overhaul of the *Teflon*, *Squirt*, and *Eveready* formats. Rather, we demonstrate that modest, tractable changes to the format of questions and answer choices will result in trademark surveys that can provide courts with far better information about the strength and meaningfulness of consumer beliefs.

A. The *Teflon* Genericism Survey Format

This section begins by reporting findings from an experiment we conducted to measure consumer uncertainty in *Teflon* surveys. It opens with a discussion of the purpose, origins, and format of *Teflon* surveys. It then turns to the design and results of our experiment. The first subsection describes how a group of randomly assigned test subjects responded when asked to take a *Teflon* survey and then rate how confident they were in the answers they gave. It then discusses how two additional groups of randomly assigned subjects responded when presented with modified *Teflon* surveys that allow beliefs to be expressed on a continuum. As detailed below, subjects expressed high levels of uncertainty about the disputed mark when exposed to each of these modifications, providing strong support for the argument that *Teflon* surveys, as currently constructed, obscure important information about the strength of consumer beliefs.

1. The Origins of the *Teflon* Survey Format

Teflon surveys are used to determine whether an asserted mark is *generic* or *distinctive of source*.¹¹¹ A term is generic if most consumers understand it not as

¹⁰⁹ See *infra* text accompanying note 115.

¹¹⁰ Many commentators have discussed the prominence of these survey formats in trademark litigation. See, e.g., Eric D. DeRosia, *Fixing Ever-Ready: Repairing and Standardizing the Traditional Survey Measure of Consumer Confusion*, 53 GA. L. REV. 613, 623 (2019) ("[T]he Ever-Ready method has become the orthodox method for measuring consumer confusion."); McCARTHY, *supra* note 1, § 32:173 (focusing on the *Eveready* and *Squirt* survey formats as "[t]wo survey formats which have become commonly used to test for confusion of source or connection" and discussing likelihood of confusion surveys in which failure to use one of these formats was "highly suspect"); McCARTHY, *supra* note 1, § 12:16 (characterizing the *Teflon* survey format as "[t]he most widely used survey format to resolve a genericness challenge").

¹¹¹ McCARTHY, *supra* note 1, § 12:16.

indicating the source of any particular product, but rather as denoting a type or category or “genus” of products. So, for example, the term “sugar” is generic for sucrose, whereas the term DOMINO is distinctive for a particular brand of sugar. Terms that function as generic labels, as “sugar” does for sucrose, do not qualify for trademark protection both because consumers do not perceive them as trademarks and because denying competitors the ability to use such terms would significantly impair competition.¹¹²

The issue of genericism arises in cases where a defendant argues that the plaintiff’s asserted mark is not protectable because it was “born generic”—i.e., that it is, and has always been, generic.¹¹³ The issue also arises in cases where a defendant offers a so-called “genericide” defense: i.e., where a defendant argues that the plaintiff’s mark may have at one time been protectable but has come to be used by consumers as a general term that describes a category of products and so no longer functions as a mark.¹¹⁴ An example of the latter sort of “acquired” genericism is the word “zipper,” which once served as a trademark for a particular producer of toothed closures but is now a generic label used to label the entire product category of toothed closures.¹¹⁵

The *Teflon* survey format was first employed in *E. I. DuPont de Nemours & Co. v. Yoshida International, Inc.*, a trademark confusion dispute between DuPont, the owner of the TEFLON family of marks, and YKK, a firm that marketed a nylon zipper under the EFLON mark.¹¹⁶ As part of its defense, YKK argued that the term TEFLON had become generic for nonstick coatings and was thus no longer protectable as a trademark.¹¹⁷ To counter this argument, DuPont introduced two surveys.¹¹⁸ The second (“Survey B”) found that 68% of respondents identified TEFLON as a brand name.¹¹⁹ That survey, which the court accepted as strong evidence in support of its conclusion that the TEFLON

¹¹² U.S. Pat. & Trademark Off. v. Booking.com B. V., 140 S. Ct. 2298, 2301–02, 2308 (2020) (“A generic name—the name of a class of products or services—is ineligible for federal trademark registration.”); *see also* McCARTHY, *supra* note 1, §§ 12:2, 12.4.

¹¹³ Stacey L. Dogan & Mark A. Lemley, *A Search-Costs Theory of Limiting Doctrines in Trademark Law*, 97 TRADEMARK REP. 1223, 1242 (2007).

¹¹⁴ *Id.*

¹¹⁵ Whitson Gordon, *How a Brand Name Becomes Generic*, N.Y. TIMES (June 24, 2019), <https://www.nytimes.com/2019/06/24/smarter-living/how-a-brand-name-becomes-generic.html> (“You may be aware of Kleenex, Velcro and ChapStick, but what about escalator? Or dumpster? Linoleum, zipper, trampoline? All of these are (or were) trademarks of companies whose products were so successful that they came to represent an entire category.”).

¹¹⁶ 393 F. Supp. 502, 505–06 (E.D.N.Y. 1975).

¹¹⁷ *Id.* at 523–24.

¹¹⁸ *Id.* at 525–26.

¹¹⁹ *Id.* at 526.

mark had not fallen prey to genericide,¹²⁰ became the model for all subsequent “*Teflon*” surveys so frequently presented in trademark litigation.¹²¹

2. *Measuring Uncertainty in Teflon Surveys*

The *Teflon* survey format is structured in two parts. The first offers what commentators have described as “essentially a mini-course in the generic versus trademark distinction” followed by a mini-test to confirm that respondents grasp the difference.¹²² This mini-test typically runs the respondent through two or three terms (such as “washing machine” and CHEVROLET) to ask whether the terms are common, generic names or brand names. After the respondents have proven that they understand the difference, the second part of the *Teflon* survey then presents the respondents with several terms, including the asserted mark at issue, and asks them to classify each as either a brand name or a common name. Importantly, the respondents are forced to choose among only three possible answers: “Brand name,” “Common name,” or “Don’t know.”¹²³

Though the second part of the *Teflon* survey format allows for a “don’t know” option, it fails to capture important information about a respondent’s degree of uncertainty. To examine this issue, we conducted a set of online experiments modeled after the *Teflon* survey used as evidence in *Snyder’s Lance, Inc. v. Frito-Lay North America, Inc.*¹²⁴ This long-running trademark dispute arose when snack food behemoth Frito Lay opposed Snyder’s Lance’s application to register the term PRETZEL CRISPS for its pretzel cracker snack.¹²⁵ Frito Lay argued that “pretzel crisps” is a generic term and therefore

¹²⁰ *Id.* at 527.

¹²¹ MCCARTHY, *supra* note 1, § 12:16.

¹²² *Id.* For example, the main survey in the *Teflon* case read:

I’d like to read 8 names to you and get you to tell me whether you think it is a brand name or a common name; by *brand* name, I mean a word like *Chevrolet* which is made by one company; by *common* name, I mean a word *like automobile* which is made by a number of different companies. So if I were to ask you, “Is *Chevrolet* a brand name or a common name?,” what would you say? Now, if I were to ask you, “Is *washing machine* a brand name or a common name?,” what would you say?

Id.

¹²³ *Id.*

¹²⁴ 542 F. Supp. 3d 371, 397–403 (W.D.N.C. 2021). The case originally began under the caption Frito-Lay N. Am., Inc. v. Princeton Vanguard, LLC, 109 U.S.P.Q.2d 1949 (T.T.A.B. 2014). Snyder’s Lance, Inc. acquired Princeton Vanguard, LLC in 2012. *Snyder’s-Lance, Inc. v. Frito-Lay N. Am., Inc.*, 991 F.3d 512, 514 n.1 (4th Cir. 2021).

¹²⁵ *Snyder’s Lance*, 542 F. Supp. 3d at 375–76.

not registerable.¹²⁶ Survey evidence played a prominent role throughout the dispute.¹²⁷

Our experiments built on the survey that E. Deborah Jay¹²⁸ developed as an expert witness for Snyder's Lance.¹²⁹ In a departure from Jay's methodology, however, we randomly assigned our 242 respondents to three groups.¹³⁰ We exposed each group to a different variation on the *Teflon* format.

a. Alternative Teflon Method A: Following Up on the Three-Answer Forced Choice Question

Respondents assigned to Group A (n = 81) proceeded through a survey nearly identical in format to Jay's, except that we added an additional follow-up question intended to probe respondents' degree of confidence in their classification of the term PRETZEL CRISPS.

Specifically, after passing the same mini-test that Jay used, Group A respondents were exposed in random order to the same six control terms that Jay used: CHEESE NIPS, MACADAMIA NUTS, ONION RINGS, GOURMET POPCORN, FLAVOR TWISTS, and SUN CHIPS. Also as Jay did, we then exposed respondents to the term PRETZEL CRISPS. Upon exposing respondents to each term, we asked: "Do you think [the term] is a brand name

¹²⁶ *Id.*

¹²⁷ See Princeton Vanguard, LLC v. Frito-Lay N. Am., Inc., 786 F.3d 960, 963–64 (Fed. Cir. 2015).

¹²⁸ E. Deborah Jay is principal of Jay Survey Strategies, LLC, where she conducts, evaluates, and testifies about litigation surveys in trademark, deceptive advertising, right of publicity, copyright, patent, wage and hour, and employee discrimination cases, among others. Jay Survey Strategies LLC, <http://www.jaysurveystrategies.com> (last visited Feb. 11, 2021).

¹²⁹ *Snyder's Lance*, 542 F. Supp. 3d at 397–98.

¹³⁰ We recruited subjects for all of the studies described in this Article through Amazon Mechanical Turk ("MTurk"). MTurk is a crowdsourcing service that allows researchers to recruit large numbers of participants for online studies. *Amazon Mechanical Turk*, AMAZON, <https://www.mturk.com> (last visited Aug. 31, 2022). People who responded to our call for subjects on MTurk were directed to an online survey we created in Qualtrics. For our *Teflon* study, 392 people responded to our call for subjects on MTurk. Following Jay's procedures, we determined whether people were eligible to participate in the study by asking the following questions: (1) "In the past 3 months, did you, personally, purchase salty snacks for you or someone else?" and (2) "In the next 3 months, do you think you, personally, will purchase salty snacks for you or someone else?" Anyone who did not answer in the affirmative to one of these questions was considered ineligible. Next, potential subjects read a set of instructions that explained the difference between a brand name and a generic name, followed by two practice questions to make sure they understood this difference. The practice questions asked subjects whether the terms BAKED TOSTITOS and TORTILLA CHIPS are brand names or generic names. Only people who answered both practice questions correctly (the first is a brand, the second generic) were permitted to participate in the study. Of the 392 people who responded to our call for subjects on MTurk, 242 were deemed eligible to participate in the full survey. These 242 subjects were then randomly assigned to one of three groups.

or a generic name?” Below the question, respondents saw three answer choices: “Generic name,” “Brand name,” and “Don’t know/Not sure.”¹³¹ After this question, we added a follow-up question that Jay did not ask. As Figure 1 shows, Group A respondents who answered “Generic name” or “Brand name” were asked: “How likely do you think it is that your answer is correct?” The answer choices were arrayed horizontally and included, “Just guessing,” “Somewhat likely correct,” “Very likely correct,” and “Definitely correct.” Group A respondents who answered “Don’t know” to the first question were not asked the follow-up question.

Figure 1
Group A Follow-up Question Format

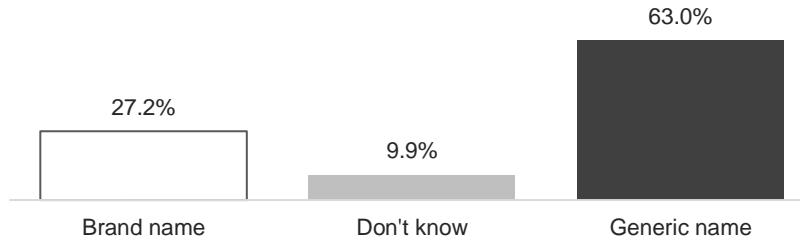
You said that you think PRETZEL CRISPS is a Brand name.

How likely do you think it is that your answer is correct?

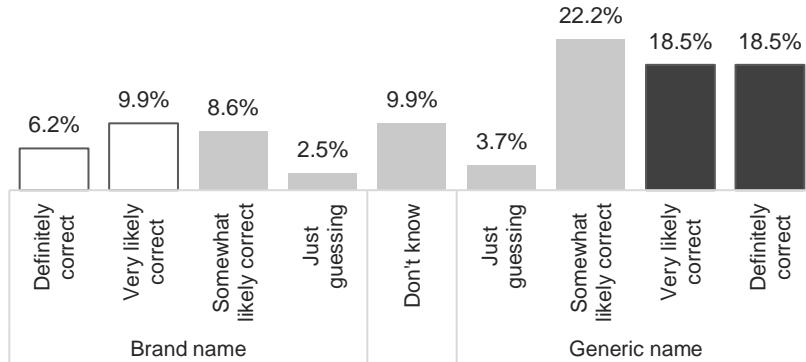
Just guessing	Somewhat likely correct	Very likely correct	Definitely correct
---------------	-------------------------	---------------------	--------------------

Figure 2 reports the distribution of Group A’s responses to the first PRETZEL CRISPS question. A clear majority of Group A respondents, 63.0%, indicated that they believed PRETZEL CRISPS to be a generic term, while just 27.2% perceived PRETZEL CRISPS as a brand name. Consistent with other *Teflon* studies, a relatively small percentage, 9.9%, responded “Don’t know/Not sure.” Taken at face value, these results strongly support a finding that the term is generic.

¹³¹ We randomized the order in which the words “brand name” and “generic name” appeared in the questions and answer choices.

Figure 2**Distribution of Group A Responses to the Standard *Teflon* Question**

The distribution of responses to the follow-up question, however, reveals substantial uncertainty hidden behind respondents' answers to the first question. Figure 3 reports the percentage of Group A subjects who gave each possible combination of answers to the first PRETZEL CRISPS question and the follow-up question. For example, the left-most bar labeled "Brand name: Definitely Correct" refers to the percentage of Group A subjects, 6.2%, who answered "brand name" to the first question and "definitely correct" to the follow-up question. The center bar represents the 9.9% of respondents who answered "Don't know" to the first question and who were thus not asked the follow-up question.

Figure 3**Distribution of Group A Responses to Both Questions**

We observe in Figure 3 that nearly half of respondents in Group A, 46.9%, expressed high levels of uncertainty about their answer to the first PRETZEL CRISPS question. The five middle bars in Figure 3, shaded light gray, represent these uncertain respondents. Remarkably, 6.2% of respondents admitted in their answer to the follow-up question that they were just guessing in their answer to the first question and an additional 30.8% of respondents indicated that they believed their answer to the first question to be only “somewhat likely correct.” Put differently, of those respondents who answered “brand name” to the first question, only 59% were confident in their answer (in that they responded “definitely correct” or “very likely correct” to the follow-up question), and of those who answered “generic name” to the first question, only 58% were confident in their answer. The remaining respondents who answered “brand name” or “generic name” were guessing or unconvincing of the belief they expressed.

Whichever way the data are parsed, it is clear that the first PRETZEL CRISPS question on its own provides, at best, limited information about the reality of consumer beliefs with respect to the term and, at worst, altogether misleading information about those beliefs.

The widespread uncertainty that Group A respondents reported may help to explain why our results on the first PRETZEL CRISPS question differed from Jay’s. We essentially re-ran Jay’s study, using the same methodology and a sample qualified using the same questions. Reassuringly, our results and Jay’s results were virtually identical on the six control terms, each of which presented a relatively easy case on which respondents would have definite beliefs.¹³² Yet with respect to PRETZEL CRISPS, while we found that only 27% of our respondents indicated that they perceived PRETZEL CRISPS as a brand name, Jay reported that 55% of her respondents did so, a statistic that supported, though not strongly, her client’s case. In essence, our study and Jay’s came to opposite

¹³² When presented with PRETZEL CRISPS, 55% of Jay’s sample and 27.2% of our sample said the term refers to a brand name. *Snyder’s Lance, Inc.*, 542 F. Supp. 3d at 398–99. Differences between the two studies were much smaller on the control terms. With regard to the three brand names, 96% of Jay’s sample and 97.5% of our sample said that SUN CHIPS is a brand name; 85% of Jay’s sample and 88.9% of our sample said that CHEESE NIPS is a brand name; and 48% of Jay’s sample and 51.9% of our sample said that FLAVOR TWISTS is a brand name. *Id.* With regard to the three generic names, 92% of Jay’s sample and 96.3% of our sample said that MACADAMIA NUTS is a generic name; 91% of Jay’s sample and 98% of our sample said that ONION RINGS is a generic name; and 72% of Jay’s sample and 86.4% of our sample said that GOURMET POPCORN is a generic name. *Id.*

conclusions about the main fact question the survey was intended to resolve.¹³³ But as we noted, when we look behind our results on the first PRETZEL CRISPS question, we find that nearly half of the respondents held weak beliefs or nonattitudes with respect to that question. We suspect that had Jay asked our follow-up question, she would have discovered that the same was true of a high proportion of individuals in her sample. We cannot prove that the instability of weakly held beliefs was the cause of the variance between our results and Jay's on the first PRETZEL CRISPS question; we cannot recreate and retest Jay's particular sample. It is revealing, however, that our results were so similar to Jay's on the control terms, which were designed, on the whole, to fall more clearly in the category of either brand name (e.g., CHEESE NIPS) or generic term (e.g., MACADAMIA NUTS). Respondents exhibited no response instability on those terms. They did so only with respect to the more difficult question of whether PRETZEL CRISPS was generic or a brand name.

Part III more thoroughly considers how a court should interpret survey data, such as the responses we recorded to the follow-up question, that indicate high levels of respondent uncertainty on the question of whether an asserted mark is perceived by consumers as generic or distinctive of source. To be sure, such data may render a survey significantly less probative of the asserted mark's distinctiveness. This may depend, however, on the particular distribution of responses and who bears the burden of persuasion on the fact issue addressed by the survey. But before pursuing these possibilities further, we explore the results of our other experiments.

b. Alternative Teflon Method B: Measuring Uncertainty on a Seven-Point Likert Scale

Respondents assigned to Group B (n = 81) proceeded through a survey very similar in format to Jay's original survey, except that for each of the seven terms, respondents were provided not with the "Generic name"/"Brand name"/"Don't know" answer choices but rather with a seven-point Likert scale as shown in Figure 4. This answer format allowed respondents to state whether they believe

¹³³ See *supra* Section I.B. In *Frito-Lay North America, Inc. v. Princeton Vanguard, LLC*, the Trademark Trial and Appeal Board criticized Jay's use in her survey's initial mini-course of the term WHEAT THINS as an example of a brand, since it is not a highly distinctive mark "and thus not a good example to participants of how to distinguish between a distinctive term and a merely well-advertised highly descriptive or even generic term." 124 U.S.P.Q.2d (BNA) 1184, 1197 (T.T.A.B. 2017). To replicate Jay's protocol, we used the same example. Thus, we cannot point to the circumstances of the mini-course to explain the difference between Jay's and our results.

the term is a brand name or a generic term and simultaneously signal their level of confidence in that belief.

Figure 4

GROUP B Question Structure: Likert Scale

Do you think PRETZEL CRISPS is a generic name or a brand name?



Figure 5 reports the distribution of Group B's responses on the PRETZEL CRISPS question. First, as a reliability check, we observe that the distribution of Group B's responses, when aggregated, is closely similar to the distribution of Group A's responses to the first PRETZEL CRISPS question (the three-answer forced-choice question). Grouping responses together, 59.3% of Group B respondents said that PRETZEL CRISPS is somewhat likely, very likely, or definitely a generic name, comparable to the 63.0% of Group A respondents who answered "Generic name" to the first PRETZEL CRISPS question. Similarly, 30.9% of Group B gave a response falling under "Brand name" (27.2% in Group A) and the proportions of "Not sure/Don't know" responses were identical across the groups at 9.9%.

Figure 5

Distribution of Group B Responses

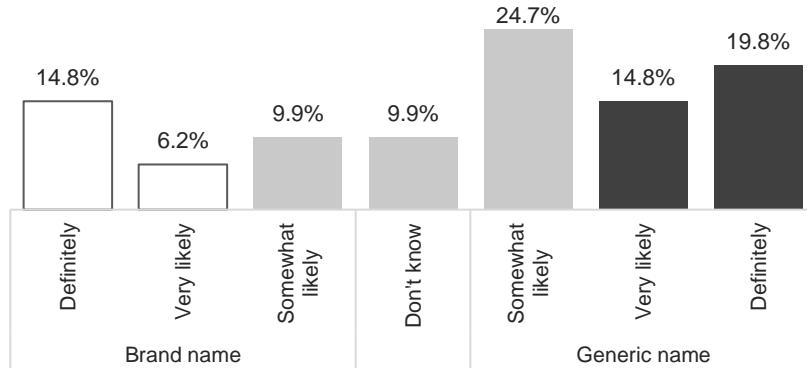


Figure 5 also shows comparably high levels of uncertainty among Group B's respondents. The three middle bars in Figure 5, shaded light gray, indicate that 44.5% of Group B subjects placed themselves at "Don't know" (9.9%), "Somewhat likely a generic name" (24.7%), or "Somewhat likely a brand name" (9.9%) on the seven-point scale. As with Group A, where 46.9% of respondents ultimately expressed high levels of uncertainty, so too with Group B we conclude that nearly half of the respondents were in or near equipoise on the question of whether PRETZEL CRISPS is a generic name or a brand name. This bottom-line finding is, of course, fundamentally different from the naive conclusion that a court might draw only from Group A's responses to the traditional *Teflon* question—specifically, the first PRETZEL CRISPS question presented to Group A suggesting that 63.0% of respondents believed the term to be generic.

We also observe that slightly higher proportions of Group B respondents expressed complete certainty in their responses than Group A respondents did. In total, 34.6% of Group B respondents answered that PRETZEL CRISPS was either "Definitely a brand name" (14.8%) or "Definitely a generic name" (19.8%). Recall that 24.7% of Group A respondents answered "Brand name: definitely correct" (6.2%) or "Generic name: definitely correct" (18.5%). Although the difference between the proportion of Group A and Group B respondents who expressed complete certainty in their answers is not statistically significant ($z = 1.38, p = 0.17$), we had expected a different result going into the experiments. Our initial expectation was that Group A respondents' initial responses to the forced choice question would produce an anchoring effect that would cause them to commit decisively to those responses in their answers to the follow-up question. But this anchoring effect did not occur (nor did it occur as we expected in our *Squirt* experiments, which we will discuss below).

c. Alternative Teflon Method C: Measuring Uncertainty Using a Continuous Scale

Respondents assigned to Group C ($n = 80$) also proceeded through a survey very similar in format to Jay's original survey, except that for each of the seven terms, respondents were provided not with Jay's "Generic name"/"Brand name"/"Don't know" answer choices but rather with a slider that ranged from 0 to 100, where 0 represented "Definitely a brand name" and 100 represented "Definitely a generic name." The slider presented respondents with a continuous

2023]

CONSUMER UNCERTAINTY IN TRADEMARK LAW

521

scale, but we placed intermediate labels on the slider between the 20 and 30 mark (“Likely a brand name”), at the 50 mark (“Don’t know/Not sure”), and between the 70 and 80 mark (“Likely a generic name”) to help guide the respondent. Figure 6 shows the format of the slider presented to Group C respondents.

Figure 6
Group C Question Structure: Slider

On a scale of 0 to 100, where 0 represents “Definitely a brand name” and 100 represents “Definitely a generic name,” do you think PRETZEL CRISPS is a brand name or a generic name?

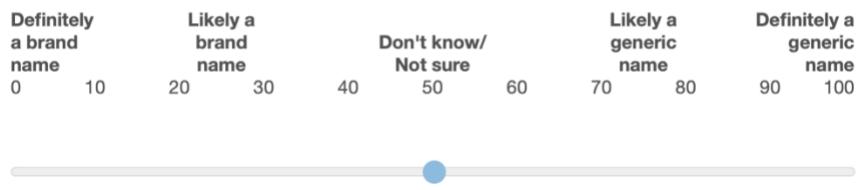


Figure 7 shows the distribution of Group C’s responses. The general patterns are similar to those we saw with Group A and Group B. The large number of intervals on the slider do not make this obvious. But, if we break the scale into three generalized categories, we find that 28.8% of Group C respondents moved the slider to a number between 0–39 (leaning brand name); 63.8% moved it to a number between 60–100 (leaning generic); and 7.4% moved it to a number between 40–59 (at or near equipoise). This distribution is remarkably similar to the distribution of Group A responses to the first PRETZEL CRISPS question (27.2%; 63.0%; 9.9%), shown in Figure 2.

Figure 7
Distribution of Group C Responses

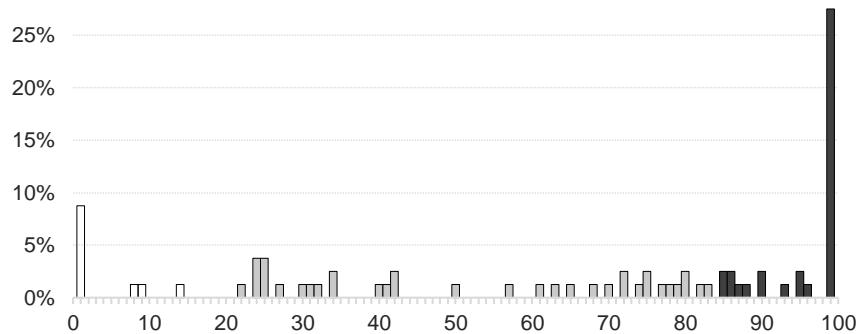


Figure 7 also provides useful information about Group C's uncertainty levels, but again, this is less obvious than in our analyses of Group A and Group B. With so many intervals on this scale, the cutoff points between more certain and less certain responses are necessarily arbitrary. But we can say with confidence that any respondents who moved the slider to 0 or 100 were expressing a high degree of certainty in their belief that PRETZEL CRISPS is either a brand name or a generic name. This included the 8.8% of respondents who moved the slider to 0 ("Definitely a brand name") and the 27.5% who moved it to 100 ("Definitely a generic name"). If we expand the range for what we consider to be a high level of certainty to include 0–14 on the left side of Figure 7 (white bars) and 86–100 on the right (black bars), 12.5% of Group C subjects believed with a high degree of certainty that PRETZEL CRISPS is a brand name and 40% believed with a high degree of certainty that it is a generic name. We are again left with just under half of Group C subjects (47.5%, shaded light gray) in what could be thought of as the range of uncertainty.

d. Evaluating the Relative Merits of Alternative Teflon Methods A, B, and C

Based on the Group A, B, and C data, we draw two clear conclusions. First, respondents to *Teflon* surveys experience high levels of uncertainty and nonattitudes which the standard *Teflon* survey format is simply not equipped to measure. Second, the scales and question types presented to Groups A, B, and C are similarly capable of measuring this uncertainty. For this reason, any of them

are superior to the dominant practice of forcing respondents to choose whether a term refers to a brand name or a generic name, even if a “Don’t know” option is explicitly or implicitly made available.

Among the three, the two-part question format shown to Group A provides the most useful information. The follow-up question gives respondents an opportunity to think about their answer to the main *Teflon* question and indicate not only how strongly they hold the belief, but also whether the belief is meaningful or just a random guess. The seven-point Likert scale (Group B) sacrifices some of this insight for efficiency. Because the Likert scale requires just one question, it demands less time and patience on the part of survey respondents than the two-part question format we showed to Group A. At the same time, we lose the ability to identify respondents who are willing to admit that they are just guessing. Finally, we are not persuaded that the 100-point slider used by Group C adds any meaningful information. In fact, it may create unnecessary confusion because cutoff points on a 100-point slider are arbitrary and up for debate. The two-part question format and the seven-point scale, on the other hand, describe each point on a continuum in language that should be clear to the respondent, the researcher, and the court.

B. The Squirt Confusion Survey Format

This section turns now to the central question in most trademark litigation: what is the *likelihood of consumer confusion*? That is, what is the likelihood that, due to the similarity between the plaintiff’s and the defendant’s marks, some appreciable proportion of relevant consumers will mistakenly believe that the defendant’s products originate with the plaintiff?¹³⁴ To determine the likelihood of consumer confusion, trademark law uses a multifactor test, one factor of which analyzes evidence of actual consumer confusion.¹³⁵ Under this factor, parties often present survey evidence of the likelihood of consumer confusion.

Survey experts have developed a number of survey formats to test for the likelihood of consumer confusion. One particular likelihood of confusion survey format is especially appropriate in situations where the plaintiff’s mark is not

¹³⁴ There are also far less common claims of so-called “reverse confusion,” in which the plaintiff complains that the defendant’s use of a similar mark causes consumers to mistakenly believe that the plaintiff’s products or services originate with the defendant. *W.W.W. Pharm. Co., Inc. v. Gillette Co.*, 984 F.2d 567, 571 (2d Cir. 1993).

¹³⁵ *E.g.*, *Polaroid Corp. v. Polarad Elecs. Corp.*, 287 F.2d 492, 495 (2d Cir. 1961) (per curiam); *Lois Sportswear, U.S.A., Inc. v. Levi Strauss & Co.*, 799 F.2d 867, 875 (2d Cir. 1986); *RESTATEMENT (THIRD) OF UNFAIR COMPETITION* § 21 (AM. L. INST. 1995).

well-known. This format is traditionally called the “*Squirt* format” after the 1980 case *Squirt Co. v. Seven-Up Co.*,¹³⁶ in which the test first came to prominence.

This section begins by discussing the origins and format of *Squirt* surveys. It then presents results from an online experiment that we designed to measure consumer uncertainty in *Squirt* surveys. Similar to our *Teflon* experiment, this experiment asked randomly assigned groups of respondents either to rate their level of confidence in answers given to the main *Squirt* question or to provide their answers to the main *Squirt* question on a seven-point Likert scale. As explained below, subjects expressed high levels of uncertainty about the relationship between the plaintiff and defendant’s products when the standard *Squirt* format was modified to allow for more nuanced responses.

1. *The Origins of the Squirt Survey Format*

The facts of *Squirt* were straightforward. SquirtCo had been using the trademark SQUIRT for a carbonated grapefruit-flavored beverage since the 1930s.¹³⁷ In 1978, Seven-Up Co. adopted the mark QUIRST for a noncarbonated lemonade beverage.¹³⁸ After a bench trial, the district court credited Squirt’s two consumer surveys and found that QUIRST infringed on SQUIRT.¹³⁹ In affirming, the Eighth Circuit characterized the surveys as “conventional in methodology and content.”¹⁴⁰

One survey randomly interviewed women twenty-five and older shopping at Zayre’s department stores in Chicago, where SQUIRT was sold but QUIRST was not.¹⁴¹ Respondents were played radio advertisements for SQUIRT and QUIRST and two other products, and were then asked “(1) ‘Do you think SQUIRT and QUIRST are put out by the same company or by different companies?’ and (2) ‘What makes you think that?’”¹⁴² Of the 152 respondents, 51 (34%) thought the two drinks were put out by the same company, 84 (55%) thought they were put out by different companies, and 17 (11%) volunteered that they did not know.¹⁴³

¹³⁶ *Squirt Co. v. Seven-Up Co.*, No. 78-375C (A), 1979 WL 25027, at *9–23 (E.D. Mo. Sept. 6, 1979) (discussing the surveys in detail), *aff’d in part, remanded in part sub nom. SquirtCo v. Seven-Up Co.*, 628 F.2d 1086, 1092 (8th Cir. 1980).

¹³⁷ *SquirtCo*, 628 F.2d at 1088.

¹³⁸ *Id.*

¹³⁹ *Squirt Co.*, 1979 WL 25027, at *23.

¹⁴⁰ *SquirtCo*, 628 F.2d at 1089 n.4, 1092.

¹⁴¹ *Id.* at 1090 n.4.

¹⁴² *Id.*

¹⁴³ *Id.*

The second survey interviewed Phoenix grocery store shoppers and was “limited to women 25 and older who had purchased soft drinks that day.”¹⁴⁴ “Because both SQUIRT soft drink and QUIRST soft drink were available [in Phoenix], no radio ads were played to establish brand awareness.”¹⁴⁵ Instead, shoppers were directly asked the same two questions as those asked in Chicago. Of the 476 respondents, 110 (23%) thought the two drinks were put out by the same company, 161 (34%) thought they were put out by different companies, and 205 (43%) volunteered that they did not know.¹⁴⁶

Interestingly, the original *Squirt* survey had a fundamental problem: it did not explicitly offer an option for responding “don’t know.” The district court never addressed this issue. The Eighth Circuit recognized that “Seven-Up’s principal attack on the survey was that the first question was dichotomous and therefore encouraged guessing by suggesting that it was improper to answer ‘I don’t know.’”¹⁴⁷ But the court did not offer any analysis of its own of the issue and apparently considered the survey to be credible.

Since the first use of the survey in the *Squirt* litigation, the *Squirt* survey format has evolved into various forms, but, at its core, it presents respondents with the plaintiff’s and the defendant’s branded products, either one after another¹⁴⁸ or side-by-side,¹⁴⁹ and then asks the respondents if they believe that the products are put out by the same company or different companies. It now also routinely includes a “don’t know” option.

Courts and commentators have criticized the *Squirt* format on the grounds that it makes the respondent “artificially aware” of the marks at issue,¹⁵⁰ but courts nevertheless often credit *Squirt* surveys.¹⁵¹

¹⁴⁴ *Id.*

¹⁴⁵ *Id.*

¹⁴⁶ *Id.*

¹⁴⁷ *Id.*

¹⁴⁸ See, e.g., *Fortune Dynamic, Inc. v. Victoria’s Secret Stores Brand Mgmt., Inc.*, 618 F.3d 1025, 1036 (9th Cir. 2010).

¹⁴⁹ See, e.g., *Bodum USA, Inc. v. A Top New Casting, Inc.*, No. 16 C 2916, 2017 WL 6626018, at *2, *6–8 (N.D. Ill. Dec. 28, 2017) (declining to exclude *Squirt* survey using “a photo array of five different French press coffee makers side by side” when this method reflected actual marketplace conditions in which products appeared together on internet shopping sites).

¹⁵⁰ *Nat’l Distillers Prods. Co. v. Refreshment Brands, Inc.*, 198 F. Supp. 2d 474, 484 (S.D.N.Y. 2002); see *id.* at 482, 484 (“[A] major flaw we discern in the [Squirt survey at issue] is that every respondent was exposed to the [plaintiff’s] product in the first room [of the two room survey], thus acquainting them with a product that they would almost certainly have been unfamiliar with otherwise, due to [the plaintiff’s] very limited distribution network and weak sales.”).

¹⁵¹ For example, in the recent case *Fortune Dynamic*, the Ninth Circuit found that the district court abused its discretion in excluding an online *Squirt* survey, particularly in light of a previous Ninth Circuit case that

2. *Measuring Uncertainty in Squirt Surveys*

To explore respondent uncertainty in *Squirt* surveys, we conducted an online experiment that exposed 201 test subjects to side-by-side images of two competing brands of flavored vodka: JOHNNY LOVE (passion fruit flavor) and PUCKER (grape flavor). As the images in Figure 8 below show, both brands employ an image of lips as part of their trade dress. The two brands were involved in the long-running dispute *JL Beverage Co., LLC v. Beam, Inc.*,¹⁵² in which defendant Jim Beam, the producer of PUCKER flavored vodka, ultimately prevailed.¹⁵³

As with our above *Teflon* experiments, our *Squirt* experiments were relatively simple. Using the facts from *JL Beverage*, we tested two variations on the standard *Squirt* survey format. We randomly assigned our 201 experimental subjects to two groups.¹⁵⁴ Each group was exposed to one of the variations.

credited a similar survey. 618 F.3d at 1030, 1037 (citing *Thane Int'l, Inc. v. Trek Bicycle Corp.*, 305 F.3d 894, 902 (9th Cir. 2002)).

¹⁵² See *JL Beverage Co. v. Beam, Inc.*, 318 F. Supp. 3d 1188, 1193 (D. Nev. 2018), *aff'd sub nom. JL Beverage Co. v. Jim Beam Brands Co.*, 815 F. App'x 110, 112, 114 (9th Cir. 2020). The Plaintiffs did not present any survey evidence at trial.

¹⁵³ See *Jim Beam*, 815 F. App'x at 112, 114.

¹⁵⁴ Similar to our *Teflon* study, we recruited subjects for our *Squirt* study through MTurk, an online crowdsourcing service. Subjects who answered our call in MTurk were then directed to an online survey where they were randomly assigned to one of two groups. *Amazon Mechanical Turk*, AMAZON, <https://www.mturk.com> (last visited Aug. 31, 2022).

Figure 8
Experimental Stimuli



a. Alternative Squirt Method A: Following Up on the Three-Answer Forced Choice Question

Survey respondents assigned to Group A (n = 102) were exposed to the side-by-side images shown in Figure 8 and were required to view them for at least five seconds before advancing. On the next pages, the images remained in view as the test subjects were asked a series of questions about the depicted products.¹⁵⁵ Among these, we asked the main *Squirt* question: “Do you believe these products are put out by the same company or by different companies?”¹⁵⁶

¹⁵⁵ At the end of the survey, all subjects were asked the same series of questions about the number of bottles of liquor they had purchased in the past twelve months and how knowledgeable they are about liquor. They were also asked basic demographic questions.

¹⁵⁶ We randomly varied the order of “by the same company” and “by different companies.”

Respondents were provided with three answer choices: “Same company,” “Different companies,” and “Don’t know/No opinion.”¹⁵⁷

Unlike in conventional *Squirt* surveys, we then asked a follow-up question to study the degree of certainty or uncertainty that Group A respondents had in their responses. Respondents who answered “Same company” to the main *Squirt* question were asked: “How likely do you think it is that these products are put out by the same company?” They were provided with four possible responses: “Just guessing,” “Somewhat likely same company,” “Very likely same company,” or “Definitely same company.” Respondents who answered “Different companies” to the main *Squirt* question were asked a similar question and provided with similar answers except that “different companies” was substituted for “the same company.”¹⁵⁸

Figure 9 shows how Group A respondents answered the main *Squirt* question, before they were asked the follow-up. Taken at face value, these basic results are unsurprising. The proportion of subjects answering “same company,” 24.5%, would be sufficient in most instances to support a finding of trademark confusion liability, although it flirts with the lower bound employed by some courts. Only 8.8% of subjects answered “Don’t know/No opinion,” which is consistent with the low percentage of survey respondents who choose that answer in surveys reported in the case law.¹⁵⁹

¹⁵⁷ We randomly varied the order of “Same company” and “Different companies” with “Don’t know/No opinion” always appearing last. Logistic regressions suggest that the ordering of the answers had no effect on the answer that subjects chose.

¹⁵⁸ The nine subjects who responded “Don’t know/No opinion” were asked, “If you had to guess, would you say these products are put out by different companies or by the same company?” They were shown five possible answers: “Very likely different companies,” “Somewhat likely different companies,” “No idea,” “Somewhat likely same company,” or “Very likely same company.” We randomly varied the order of these answers so that they proceeded from “Very likely different companies” to “Very likely same company” or vice-versa. Six responded “Somewhat likely same company,” two responded “No idea,” and one responded “Somewhat likely different companies.” *See infra* Figure 10.

¹⁵⁹ *See supra* note 16 and accompanying text; *see infra* Figure 9.

2023]

CONSUMER UNCERTAINTY IN TRADEMARK LAW

529

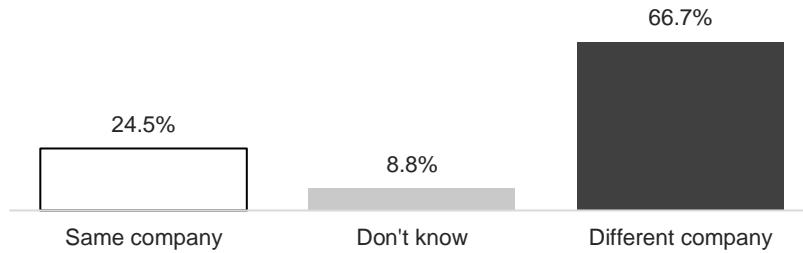
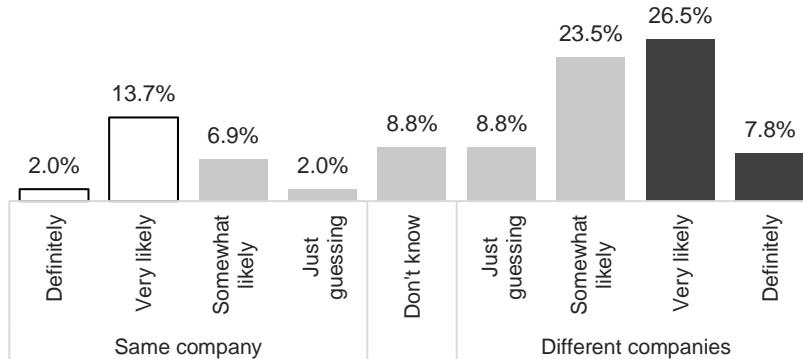
Figure 9**Group A Distribution of Responses to Main *Squirt* Question**

Figure 10 shows the percentage of Group A subjects who gave each possible combination of answers to the main *Squirt* question and the certainty follow-up question. For example, the left-most bar labeled “Same company: Definitely” refers to the percentage of Group A subjects who answered “Same company” to the main *Squirt* question and “Definitely same company” to the certainty follow-up. The middle bar represents those subjects who said “Don’t know/No opinion” to the main *Squirt* question.

Figure 10**Group A Distribution of Responses to the Main *Squirt* Question and Certainty Follow-Up Question**

The data reported in Figure 10 are remarkable. As we had in tests of the *Teflon* survey, we expected that respondents' answers to the main *Squirt* question would produce an anchoring effect that would cause them to commit decisively to those answers in their responses to the follow-up question. Again, this anchoring effect did not occur. In fact, while 24.5% of Group A respondents initially indicated that they believed that the products were put out by the same company, only 2% of Group A respondents expressed full certainty in that belief by selecting "Definitely same company" in response to the follow-up question. And, while 91.2% of Group A respondents expressed a belief in response to the first question, significant proportions then admitted that they were actually "just guessing" (10.8%) or that it was only "somewhat likely" that the products were put out by the same company (6.9%) or by different companies (23.5%). Ultimately, as indicated by the light gray bars in Figure 10, fully 50% of Group A respondents (51 out of 102) indicated a substantial degree of uncertainty about their answers. Once we take uncertainty into account, we are left with only 15.7% of subjects who answered "same company" with any certainty (as indicated by the white bars)—a number substantially lower than the 24.5% of subjects who appeared to be confused in the initial survey and below the level that most courts would consider sufficient to support a finding of consumer confusion.

b. Alternative Squirt Method B: Measuring Uncertainty on a Seven-Point Likert Scale

Survey respondents assigned to Group B ($n = 99$) were asked the same main *Squirt* question that was asked of Group A's subjects. The difference was that Group B subjects were provided with seven possible responses arrayed horizontally across the screen, as shown in Figure 11.¹⁶⁰

Figure 11

Group B Possible Responses to Main *Squirt* Question

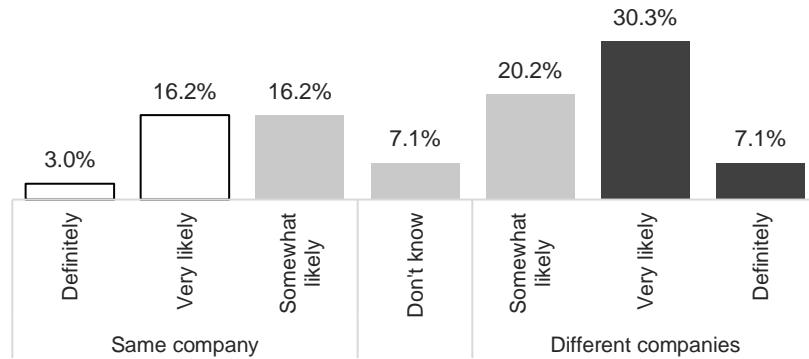
Do you believe these products are put out by the same company or by different companies?

Definitely same company	Very likely same company	Somewhat likely same company	Don't know/No opinion	Somewhat likely different companies	Very likely different companies	Definitely different companies
-------------------------	--------------------------	------------------------------	-----------------------	-------------------------------------	---------------------------------	--------------------------------

¹⁶⁰ We randomly varied the order of these answers so that they proceeded from "Definitely same company" to "Definitely different companies" or vice-versa. Logistic regressions suggest that the ordering of the answers had no effect on the answers that subjects choose.

Figure 12 shows the distribution of responses to the seven-point scale. When aggregated into three bins, Group B's distribution of responses is roughly similar to Group A's distribution of responses to the main *Squirt* question as shown in Figure 9.¹⁶¹ But, just as we saw with the uncertainty follow-up question asked of Group A respondents, Group B's seven-point Likert scale yielded a more nuanced picture of the strength of respondents' beliefs. On aggregate, 35.4% of Group B respondents signaled that they believed the products were put out by the same company (the three leftmost bars in the figure). But only 3% indicated that they believed that these products were *definitely* put out by the same company, which is comparable to the 2.0% of Group A respondents who provided the same response. In total, only 19.2% of Group B respondents indicated with a high degree of certainty that they believed the two vodka products were put out by the same company.

Figure 12
Group B Distribution of Responses to Main *Squirt* Question



At the same time, a plurality of Group B respondents expressed high levels of uncertainty. In addition to the 7.1% of respondents who selected the middle

¹⁶¹ See *infra* Figure 12. For example, while 24.5% of Group A respondents answered "Same company," 35.4% of Group B respondents answered "Definitely same company," "Very likely same company," or "Somewhat likely same company." While 66.7% of Group A answered "Different companies," 57.6% of Group B respondents provided a comparable answer. The aggregated percentages suggest that Group B subjects were more likely to believe that the products were put out by the same company (34.5% for Group B vs. 24.5% for Group A). This difference, however, is not statistically significant ($p = 0.093$).

option and admitted to having no opinion, another 36.4% placed themselves just on either side of that option when they selected “Somewhat likely same company” or “Somewhat likely different companies.” In total, 43.5% of Group B respondents appeared undecided or uncertain on the question of whether the two products were put out by the same or different companies. This is comparable to the 50% of Group A subjects who said “Don’t know,” “Just guessing,” or “Somewhat likely” to the main *Squirt* question and certainty follow-up.

As we saw with the conventional *Teflon* survey format, the conventional *Squirt* survey format fails to detect significant levels of consumer uncertainty and, in fact, conceals it behind the kind of simple and misleading bottom-line results shown in Figure 9. Minor, easily implemented modifications to the *Squirt* format reveal high proportions of respondents with weakly-held beliefs or nonattitudes. We explained above how basic social science understandings of consumer behavior have long recognized that such weakly held beliefs are unlikely to impact consumer behavior.¹⁶² Yet, conventional survey formats disguise them as strong evidence of consumer perception in the marketplace.

C. The Eveready Confusion Survey Format

We now turn to what is probably the leading survey format in trademark law, the *Eveready* survey format for the likelihood of consumer confusion. We begin by tracing the origins of the *Eveready* survey. Then, as we did earlier for the *Teflon* and *Squirt* surveys, we present results from an online study we conducted to measure uncertainty in *Eveready* responses. As detailed below, our study replicated and modified an *Eveready* survey that was originally administered in a trademark dispute brought by the makers of Budweiser beer. Our modification allowed respondents to rate their level of confidence in answers given to the main *Eveready* questions. Providing perhaps the strongest evidence yet for our argument that trademark surveys obscure consumer uncertainty, the vast majority of respondents to our *Eveready* survey admitted that the beliefs they expressed were either just guesses or weakly held.

1. The Origins of the Eveready Survey Format

As explained above, the Seventh Circuit’s emphatic endorsement of the plaintiff’s survey evidence in *Union Carbide Corp. v. Ever-Ready Inc.*¹⁶³

¹⁶² See *supra* Section I.B.

¹⁶³ 531 F.2d 366, 385 (7th Cir. 1976).

dramatically enhanced the reputation of consumer survey evidence in trademark litigation.¹⁶⁴ Of the two surveys that Union Carbide produced, the lamp survey proved to be most influential. It showed respondents the defendant's lamp bearing the EVER-READY mark and then asked: "Who do you think puts out the lamp shown here?"; "What makes you think so?"; and "Please name any other products put out by the same concern which you think puts out the lamps shown here."¹⁶⁵

The *Eveready* survey format has since evolved into a substantially more elaborate script of questions.¹⁶⁶ In essence, after being shown the defendant's product bearing the defendant's trademark, respondents are asked three sets of questions, each designed to elicit responses related to different types of potentially actionable consumer confusion. The first set of questions asks what company puts out the defendant's product or service and what makes the respondents say so.¹⁶⁷ This question tests for source confusion. The second set asks whether the company that puts out the defendant's product or service is *connected or affiliated* with any other company, and what makes the respondent say so.¹⁶⁸ And the third asks whether defendant's product or service is *approved or sponsored* by any other company, and what makes the respondent says so.¹⁶⁹ The second and third sets of questions are designed to elicit responses relevant to potential "sponsorship or affiliation confusion"—i.e., the notion that, although consumers might not believe that defendant's product or service actually originates with the plaintiff, they might nonetheless believe that the defendant is connected to the plaintiff either through ownership (connection/affiliation) or through a licensing, promotion, or marketing agreement (approval/sponsorship).¹⁷⁰

We suspect that asking consumers to report their impressions about *legal relationships* like sponsorship or affiliation—matters about which consumers often are likely to have only the vaguest impressions—is likely to produce even more uncertainty than questions about the origin of products or services. We

¹⁶⁴ See *supra* notes 97–107 and accompanying text.

¹⁶⁵ Union Carbide Corp. v. Ever-Ready Inc., 392 F. Supp. 280, 292 (N.D. Ill. 1975).

¹⁶⁶ See Starbucks U.S. Brands, LLC v. Ruben, 78 U.S.P.Q.2d 1741, 1753 (T.T.A.B. 2006) (noting that the original *Eveready* survey did not ask questions concerning sponsorship, affiliation, permission, and approval, but recognizing that courts and commentators have since come to believe it appropriate to ask such questions in the *Eveready* format).

¹⁶⁷ See MCCARTHY, *supra* note 1, § 32:174.

¹⁶⁸ See *id.*

¹⁶⁹ See *id.*

¹⁷⁰ See Int'l Info. Sys. Sec. Certification Consortium, Inc. v. Sec. Univ., LLC, 823 F.3d 153, 161–63 (2d Cir. 2016) (holding that trademark law prohibits confusion as to sponsorship, affiliation, or connection).

expect this uncertainty to be further compounded by the open-ended nature of the *Eveready* format. *Eveready* surveys expose respondents only to the defendant's mark. The respondents must then think of and state their own answers. In contrast to *Squirt* and *Teflon* surveys, respondents to *Eveready* surveys are not given any options to choose from. This is why a plaintiff will typically avoid the *Eveready* survey format if its mark is not widely known and will use instead the *Squirt* format, which exposes the respondents to both parties' marks. By contrast, a plaintiff with a very well-known mark will strongly prefer an *Eveready* survey. When exposed by the *Eveready* format to the defendant's similar mark, respondents familiar with the plaintiff's mark may simply be reminded of that mark and name it or the plaintiff.

2. *Measuring Uncertainty in the Eveready Surveys*

To measure uncertainty in *Eveready* surveys, we built upon a survey presented in *Anheuser-Busch, LLC v. Innvopak Systems Pty Ltd.*¹⁷¹ In this Trademark Trial and Appeal Board¹⁷² dispute, Anheuser-Busch opposed Innvopak Systems' application to register the trademark WINEBUD. Anheuser-Busch argued that consumers are likely to confuse the "bud" in WINEBUD with the BUDWEISER and BUD LIGHT beer brands.¹⁷³ In support of this position, Edward Blair,¹⁷⁴ an expert for Anheuser-Busch, presented results from an *Eveready* survey he conducted with 400 respondents. Blair reported that 24% of survey respondents confused WINEBUD with Anheuser-Busch or one of its brands.¹⁷⁵ The Board found Blair's results convincing and ruled that the WINEBUD mark was likely to cause confusion with Anheuser-Busch marks.¹⁷⁶

¹⁷¹ *Anheuser-Busch, LLC v. Innvopak Sys. Pty Ltd.*, 115 U.S.P.Q.2d 1816 (T.T.A.B. 2015).

¹⁷² The Trademark Trial and Appeal Board ("TTAB") is an administrative board within the United States Patent and Trademark Office ("USPTO") that hears and decides adversary proceedings, including trademark oppositions (initiated when a party opposes a mark prior to final grant and after publication in the Official Gazette) and cancellations (initiated when a party seeks to cancel an existing registration). *See Trademark Trial and Appeal Board*, USPTO, <https://www.uspto.gov/trademarks/ttab> (last visited Jan. 10, 2023). The TTAB also handles trademark interference and concurrent use proceedings, as well as appeals of final refusals issued by USPTO Trademark Examining Attorneys within the course of the prosecution of trademark applications. *Id.*

¹⁷³ *Anheuser-Busch*, 115 U.S.P.Q.2d at 1816.

¹⁷⁴ Blair is a professor at the University of Houston's Bauer College of Business. *See* Edward Blair, UNIV. OF HOUS. C.T. BAUER COLL. OF BUS., <https://www.bauer.uh.edu/search/directory/profile.asp?firstname=Edward&lastname=Blair> (last visited Dec. 12, 2022).

¹⁷⁵ *Anheuser-Busch*, 115 U.S.P.Q.2d at 1828.

¹⁷⁶ *Id.* at 1832.

To calculate confusion, Blair told respondents that a new brand of wine products called WINEBUD may be introduced to the market soon, and then asked the following open-ended questions:

- In your opinion, what company puts out these wine products?
- In your opinion, is the company that puts out these wine products affiliated or connected with any other company? (If YES) What is that other company?
- In your opinion, are these wine products approved or sponsored by any other company? (If YES) What is that other company?¹⁷⁷

Blair classified respondents as confused if they referred to Anheuser-Busch or its brands in response to at least one of these three questions.¹⁷⁸

We conducted an online survey of 1,034 respondents that was modeled after Blair's. Following each of the three open-ended questions, however, we added a new question that asked respondents to indicate how likely they thought it was that the answer they typed was correct.¹⁷⁹ Respondents were provided with four answer choices arrayed across the screen horizontally: (1) "No idea—just guessing," (2) "Somewhat likely correct," (3) "Very likely correct," or (4) "Definitely correct."

Our aggregated results on the three open-ended questions look similar to Blair's. In our survey, 165 out of 1,034 respondents (16.0%) typed Anheuser-Busch or one of its brands when asked what company puts out WINEBUD. An additional 18 respondents stated that WINEBUD is affiliated or connected with or approved or sponsored by Anheuser-Busch or its brands. Thus, according to the empirical definition of confusion used by Blair in his testimony, 183 of 1,034 respondents

¹⁷⁷ See Applicant's Brief at 11–12, Anheuser-Busch, Inc. v. Innopak Systems Pty. Ltd., 115 U.S.P.Q.2d 1816 (2015) (No. 91194148); *Anheuser-Busch*, 115 U.S.P.Q. at 1828 & n.15.

¹⁷⁸ *Id.* at 11 (noting that Blair counted as confused any subject "who said that these wine products are put out by, affiliated, or connected with, or approved or sponsored by Anheuser-Busch or its brands").

¹⁷⁹ As with our *Teflon* and *Squirt* studies, we recruited subjects for our *Eveready* study through the online crowdsourcing service MTurk. 1,252 people answered our call for subjects on MTurk. 188 people were deemed ineligible because they reported that they had not purchased at least one bottle of wine in the previous twelve months. We also omitted thirty subjects because it was clear from their responses that they had not completed the survey in good faith (e.g., they had typed numbers instead of words, pasted in long strings of text, or typed the case citation, presumably after searching for WINEBUD on the internet). After omitting ineligible subjects and invalid responses, we were left with 1,034 subjects. Our survey asked three open-ended questions similar to those that Blair asked: "In your opinion, what company puts out WINEBUD?"; "In your opinion, is the company that puts out WINEBUD affiliated or connected with any other company?" (If YES) "What is that other company?"; and "In your opinion, is WINEBUD approved or sponsored by any other company?" (If YES) "What is that other company?" Following each open-ended question, we reminded subjects of the answer they had just typed and asked, "How likely do you think it is that your answer is correct?"

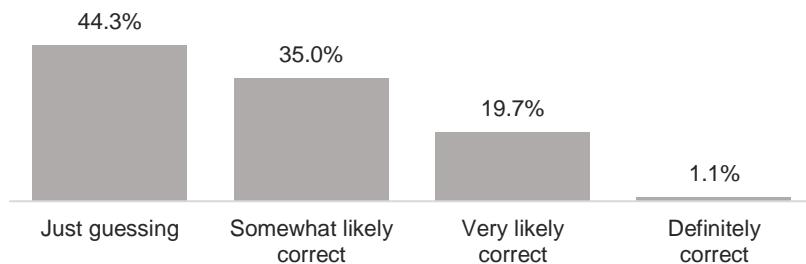
(17.7%) were confused in our study. Though the 17.7% rate of confusion we observed is not as high as the 24% that Blair reported in his study, some courts have accepted even lower proportions as evidence of a likelihood of confusion.¹⁸⁰ For this reason, taken at face value, even 17.7% would be a troubling result for at least some defendants, particularly those seeking summary judgment.

A dramatically different view emerges from the distribution of the respondents' answers to the follow-up questions. As Figure 13 shows, of the 183 respondents who said WINEBUD is put out by, affiliated or connected with, and/or approved or sponsored by Anheuser-Busch or its brands, 81 (44.3%) indicated "No idea—just guessing" when asked how likely they think it is that their answer is correct.¹⁸¹ An additional 64 of the 183 respondents (35%) followed their response by indicating that they thought their answer was only "Somewhat likely correct." In sum, 79.3% of respondents who mentioned Anheuser-Busch or its brands expressed high levels of uncertainty about their answers. Meanwhile, only 2 of 183 respondents (1.1%) indicated that they believed their answer to be "Definitely correct" and 36 of 183 respondents (19.7%) indicated "Very likely correct." Ultimately, of the 1,034 respondents surveyed, just 38 (3.7%) referred to Anheuser-Busch or its brands in a response to an open-ended question and followed that response by indicating that they had a high degree of confidence in their answer.

Figure 13

Uncertainty Among Subjects Who Mentioned Anheuser-Busch or its Brands

(n = 183)



¹⁸⁰ See *supra* text accompanying notes 39–40.

¹⁸¹ Forty-five respondents typed Anheuser-Busch or its brands in response to two of the open-ended questions and seven respondents typed Anheuser-Busch or its brands in response to all three questions. In these instances, we only counted the follow-up response given after the first time a respondent cited Anheuser-Busch or its brands. In 165 of 183 instances, respondents cited Anheuser-Busch or its brands after the first open-ended question, "In your opinion, what company puts out WINEBUD?"

Our results suggest that the *Eveready* survey format may create substantial demand effects; i.e., the “bias that occurs when participants infer the purpose of an experiment and respond so as to help confirm a researcher’s hypothesis.”¹⁸² Many respondents who have no actual beliefs regarding the survey questions may report one anyway—perhaps in a bid to be helpful to the researcher, especially because the respondents are not actually being asked to purchase anything. Relatedly, the structure of the *Eveready* survey raises the likelihood that the survey will produce responses reflecting nonattitudes. Respondents may name a plaintiff’s well-known brand simply because the defendant’s mark reminds the respondent of that brand, and not because the respondent actually believes with any certainty that the brands come from the same source. Moreover, our results show that these effects will go undetected unless the survey probes for uncertainty.¹⁸³ This possibility is addressed further below. More generally, our results show that trademark litigants should be wary of any *Eveready* survey that fails to test for uncertainty. For example, given that Blair’s respondents in his original survey were likely also experiencing some degree of uncertainty, we should question whether Blair’s original bottom-line finding of 24% confusion was helpful to the Board.

III. IMPLICATIONS

Our experiments reveal that consumers experience varying degrees of uncertainty in assessing whether a mark is generic or distinctive, or whether two similarly-branded products originate from the same source. Current standard trademark survey formats fail to register these degrees of uncertainty. This failure raises a number of normative questions that we consider here:

¹⁸² Jonathan Mummolo & Eric Peterson, *Demand Effects in Survey Experiments: An Empirical Assessment*, 113 AM. POL. SCI. REV. 517, 517 (2019). For further discussion of demand effects in survey experiments, see generally *id.*

¹⁸³ For example, *Eveready* surveys sometimes frame the third “approval or sponsorship” question to ask the respondent to state whether the defendant “needed to get permission” for their use or “did not need to get permission.” *See, e.g.*, Report of Jacob Jacoby, Ph.D. at 12, Juicy Couture, Inc. v. L’Oreal USA, Inc. (S.D.N.Y. 2005) (No. 04 CIV 7203) (submitted on behalf of defendant L’Oreal) (emphasis omitted) (report on file with authors); *see also* Report of Jacob Jacoby, Ph.D at 4, Smith v. Wal-Mart Stores, Inc. (N.D. Ga. 2005) (No. 1:06-cv-526-TCB) (report on file with authors). This question explicitly asks for a *legal* conclusion—i.e., whether the defendant’s trademark use requires permission of the senior user. The beliefs of lay survey respondents about the content of legal rules are especially likely to be weakly held, and, thus, also especially likely to respond to respondents’ perceptions about the answer that the survey-taker would expect or prefer.

- 1) Should trademark law intervene to protect consumers who believe that a mistaken proposition is only somewhat likely to be true?
- 2) Should that intervention be undertaken without regard to whether a countervailing proportion of consumers correctly recognize with complete certainty that the proposition is in fact false?¹⁸⁴
- 3) Is it efficient to respond to confused consumers with weakly held incorrect beliefs by removing the “confusing” use from the marketplace?
- 4) Is it efficient if doing so means disrupting the strongly held correct beliefs of non-confused consumers? If it forces junior mark holders to incur the costs of constructing new product/mark associations? If it requires non-confused consumers to incur the costs of internalizing those new associations?

There are also practical questions:

- 5) If finders of fact in trademark litigation should take into account consumer uncertainty, then how exactly should they do so?
- 6) What is the appropriate threshold of belief strength that would satisfy a plaintiff’s *prima facie* case, meet its burden of persuasion, or justify a blanket injunction of the defendant’s conduct?

The following section turns first to the normative questions.

A. *Consumer Uncertainty and the Expansion of Trademark Rights*

Many courts and scholars have observed that trademark rights have expanded dramatically over the past century.¹⁸⁵ Plaintiffs have enjoyed a regime in which ever more things qualify for trademark protection and that protection has grown ever broader in scope.¹⁸⁶ We believe that current trademark law survey methods have substantially contributed to this expansion. In particular, by concealing consumer uncertainty, they have intensified a “circularity”

¹⁸⁴ The position that trademark law should show solicitude for consumers who do not correctly identify the mark would represent a strong form of paternalism that greatly benefits trademark plaintiffs. But it would also penalize consumers who have strongly held true beliefs, who will be forced to do the work necessary to build new source associations if the law forces the junior user to adopt a different mark. It might also interfere with trademark law’s mission to provide consumers with “shortcuts” that make competitive markets more efficient.

¹⁸⁵ See, e.g., Penn. State Univ. v. Vintage Brand, LLC, No. 21 Civ. 01091, 2022 WL 2760233, at *4 (M.D. Pa. July 14, 2022) (surveying the expansion of trademark rights); Mark A. Lemley, *The Modern Lanham Act and the Death of Common Sense*, 108 YALE L.J. 1687, 1697–1713 (1999) (discussing the “[e]xpanding [b]oundaries of [t]rademark [r]ights”).

¹⁸⁶ See Beebe, *supra* note 5, at 2067.

dynamic already present in trademark law.¹⁸⁷ Enforcement of trademark rights causes consumers to be lulled into complacency, leading them to approach questions of product source less critically and to rely on trademarks more reflexively, which in turn leads to calls for even more stringent enforcement of trademark rights.

This is not because the current methods are bound to err in only one direction. Survey methods that elide uncertainty may either overstate or understate the extent to which respondents identify a term as distinctive, or regard the similarity between two marks as suggesting a common source, depending on the particular case.

Indeed, we see errors in both directions in the experiments in this Article. Taking account of belief strength in our *Teflon* survey¹⁸⁸ showed that the conventional form of the survey overstated respondents' beliefs in the term's genericism (a form of error that would benefit the defendant in litigation). On the other hand, our *Squirt* and *Eveready* experiments showed error in the other direction¹⁸⁹—the conventional form of these surveys tended to overstate the degree of consumer confusion (an error which helps plaintiffs).

We should not be reassured, however, by the apparent bidirectionality of the conventional surveys' error, for when the surveys interact with applicable standards of proof, the error may suddenly become directional. Plaintiffs ordinarily bear the burden of proof on both distinctiveness and the likelihood of confusion.¹⁹⁰ Plaintiffs are unlikely, in the main, to rely on surveys that are not helpful. Rather, they will tend to rely only on surveys that produce bottom-line distinctiveness or confusion numbers that advance those claims. For these surveys, it is likely that submerging consumer uncertainty is likely to help plaintiffs disproportionately.

Importantly, where consumer uncertainty works in plaintiffs' direction—for example, in a likelihood of confusion survey where the “same company” respondents are highly certain and the “different company” respondents less so—surfacing that uncertainty is likely to be only modestly helpful so long as the bottom-line confusion number exceeds the courts' undemanding thresholds.

¹⁸⁷ *Id.*

¹⁸⁸ See discussion *supra* Section II.A.

¹⁸⁹ See discussion *supra* Sections II.B, II.C.

¹⁹⁰ The defendant bears the burden of proving that a mark is generic when the mark has been registered and in continuous use for at least five consecutive years and other requirements have been met allowing the mark owner to apply for and receive “incontestable” status under Section 33 of the Lanham Act, 15 U.S.C. § 1115(b).

Where consumer uncertainty is likely to have a real effect is where it tends to undercut the strength of the plaintiff's evidence—where, in other words, substantial uncertainty pervades the beliefs of respondents who consider a term distinctive, or who believe that defendant's product originates with the plaintiff.

Given this asymmetry, the failure of trademark survey methods to account for consumer uncertainty works, over time, to expand the scope of trademark rights. In the case of distinctiveness surveys, this expansion is direct—failing to surface consumer uncertainty is likely, over time, to lead to more terms being protected as trademarks than would otherwise be the case. And the increased prevalence of marks then feeds through to shape the development of consumer perception.

That is, if the law protects certain designations as trademarks (for example, a yellow color of adhesive note pads),¹⁹¹ over time consumers will come to expect that similar designations (a red outsole on shoes)¹⁹² also function as trademarks. The bias that we observe in current trademark distinctiveness surveys is not merely affecting the outcome of individual disputes. It is likely to have systemic effects over time.

In the case of trademark confusion claims, the connection is similar, if perhaps a bit more complex. A feedback loop connects trademark litigation to consumer behavior and perception in the marketplace: consumer beliefs largely determine legal outcomes in trademark law, but those legal outcomes can, in turn, influence consumer beliefs. As courts disallow conduct that creates even weak levels of confusion within certain consumers, consumers are increasingly confronted with a marketplace in which these sorts of uses are barred. Since vigilance is costly, consumers operating in a marketplace that trademark law has made “safe” for them are likely over time to reduce their vigilance.

This is the circularity alluded to earlier: a dynamic internal to trademark law which, over time, reinforces the need for aggressive trademark enforcement.¹⁹³

¹⁹¹ See The mark consists of the color canary yellow used over the entire surface of the goods, Registration No. 2,390,667 (“The mark consists of the color canary yellow used over the entire surface of the goods.”); see also David Lazarus, *3M Not Happy with Microsoft's Post-It Emulation*, WIRED (Jan. 9, 1997), <https://www.wired.com/1997/01/3m-not-happy-with-microsofts-post-it-emulation/#:~:text=Minnesota%20Mining%20%26%20Manufacturing%2C%20better%20known,in%20damages%20and%20legal%20costs.>

¹⁹² See Christian Louboutin S.A. v. Yves Saint Laurent Am. Holding, Inc., 696 F.3d 206, 228 (2d Cir. 2012) (holding that the plaintiff's “Red Sole Mark” consisting of a “red lacquered outsole that contrasts with the color of the adjoining ‘upper’” is protectable as a trademark).

¹⁹³ See Beebe, *supra* note 5, at 2067 (discussing the process by which “lower levels of search sophistication have led the law to grant a broader scope of protection” to trademarks).

Our failure to account for uncertainty, which favors plaintiffs on balance, strengthens this dynamic.

Finally, our current failure to inquire into the strength of individual beliefs poses a broader threat to trademark's *raison d'être*, which is to promote competition in markets by providing consumers accurate information about product source. Current survey methods expose trademark courts to the risk of producing numerous false positives—i.e., cases in which plaintiffs prevail even where the evidence of either or both of distinctiveness or confusion is weak. And the error costs of false positives in the trademark context are far from trivial. When trademark courts find incorrectly that an asserted mark has acquired distinctiveness, the effect is to withdraw from competitive use a *descriptive* term; the cases involving that category of terms are precisely the ones in which proof of acquired distinctiveness is required—i.e., unlike so-called “inherently” distinctive fanciful, arbitrary, or suggestive marks, merely descriptive marks must be proven to have *acquired* distinctiveness through use.

The result may be to hobble competitors, who are deprived of free use of a term that describes their product.¹⁹⁴ And because the term does not in reality serve as a source identifier, the harm of withdrawing a descriptor from full competitive use is not counterbalanced by the benefits to competition that accrue from the protection of terms that consumers treat as source identifiers. If consumers do not, in reality, rely on the term to identify source, either because they lack a belief about the term's distinctiveness entirely or because that belief is so weakly held that consumers do not *act on it*, then we will not benefit from the reduction in information costs that we expect trademarks to provide.

The same is true where trademark courts incorrectly find consumer confusion based on survey evidence that assesses belief but not belief strength. If consumers are not in fact confused, or if their beliefs about the common source of the plaintiff's and the defendant's products are so weakly held that individuals are unlikely to act upon them in real-world markets, then trademark courts are

¹⁹⁴ Trademark's doctrine of descriptive fair use generally permits competitors to make descriptive uses of terms protected as marks. *See MCCARTHY, supra* note 1, § 11:45. That said, the boundaries of what constitutes a descriptive versus a source-indicating use are unclear and depend entirely on context, are likely to be contested in particular cases, and cannot be clarified without expensive litigation that is risky for would-be descriptive users. As a consequence, we cannot rely on trademark's descriptive fair use defense to dispel the harm caused by court's false positives on the question of distinctiveness, although the availability of the defense certain reduces that harm somewhat.

depriving non-confused consumers of a source identifier *that they use to correctly identify the source of products.*¹⁹⁵

That is, trademark courts are depriving non-confused consumers of valid information regarding source. And, again, there is no counterbalancing trademark benefit: if consumers are not confused, or if mistaken beliefs about source are weakly held and therefore unlikely to be acted upon, then legal intervention does little to dispel either confusion or “free-riding” on the senior user’s goodwill, for neither of those evils is present. Legal intervention in such cases is all social cost, with no benefit other than the private benefit to the senior mark owner of reduced competition—a private benefit *which is itself a component of the social cost imposed by the mistaken (false-positive) finding of “confusion,”* and is therefore not offsetting.

If courts produce enough of these false positives respecting either a mark’s distinctiveness or the likelihood of confusion, we might see substantial impairment of trademark law’s mission to provide a pro-competitive source of truthful market information. We might also see trademarks impose barriers to competition, barriers that create leverage that senior users—which tend to be incumbent firms—may use to obtain and to preserve market power that trademark law was never meant to provide or protect.

For all of these reasons, if trademark courts are interested in survey evidence that helps trademark law discharge its mission by actually illuminating whether consumers in the real world are likely to act on their belief that a term is distinctive, or that two products come from the same source, then they should be concerned not just about whether consumers hold those beliefs, but whether they hold them *with some threshold level of certainty.*

B. Incorporating Consumer Uncertainty into Trademark Practice

What, then, should courts require as the appropriate threshold level of certainty when analyzing trademark survey evidence? Courts should find that an asserted mark is in fact perceived by consumers as a designation of source only when a threshold proportion of consumers believe it to be at least *substantially*

¹⁹⁵ Michael Grynberg, *The Road Not Taken: Initial Interest Confusion, Consumer Search Costs, and the Challenge of the Internet*, 28 SEATTLE U. L. REV. 97, 99 (2004). In an important recent article, Daniel Gervais and Julie Latsko expand on the problems that arise when trademark law fails to take into account the interests of *non-confused* consumers. See Daniel Gervais and Julie M. Latsko, *Who Cares About the 85 Percent? Reconsidering Survey Evidence of Online Confusion in Trademark Cases*, 96 J. PAT. TRADEMARK OFF. SOC’Y 265, 270 (2014) (“[I]f the likelihood of consumer confusion of a small minority is minimal but an injunction would impose a significant loss/cost on other consumers[,] . . . then the injunction should not issue.”).

likely that the mark is a designation of source. Similarly, courts could find infringement only when a threshold proportion of consumers believe it to be at least *substantially likely* that the defendant's products originate in the plaintiff.

More specifically, when analyzing the distribution of uncertainty in any surveyed sample of consumers, courts could adopt a “top two box” approach that focuses on the “definitely” and “very likely” categories in a Likert scale. This approach is preferable for reasons grounded both in social science and the mechanics of trademark litigation.

As a matter of social science, consumer behavior research has long studied methods of aggregating responses to five-point (as well as seven-, nine-, and eleven-point) scales to predict actual consumer behavior. Researchers are particularly concerned that their surveys may produce demand effects.¹⁹⁶ One strategy to address this problem is to apply a weighting scheme where the weights decrease as the stated likelihood of purchase decreases. But research has shown that weighting schemes are not absolute and vary with the type of product.¹⁹⁷ This can make such schemes difficult to generalize—and in a litigation context subject to expert manipulation.

More commonly, consumer behavior researchers will take the percentage of respondents in the top or top two boxes of a multipoint scale as indicators of purchase intention.¹⁹⁸ Various empirical studies of consumer intentions and consumer purchasing conduct have confirmed a “threshold effect” that is consistent with this top box or top two boxes approach.¹⁹⁹

From the perspective of the mechanics of trademark litigation, the legal standard in trademark cases (as in virtually all civil litigation) is preponderance of the evidence—i.e., that consumer confusion is more likely than not. And once we redesign trademark surveys to take belief strength into account, it is only the top two points on the Likert scale that represent a belief arguably strong enough to indicate that the proposition is more likely true than not. Beliefs of this

¹⁹⁶ On the meaning of “demand effects,” see *supra* text accompanying note 182.

¹⁹⁷ See generally Vicki Morwitz et al., *supra* note 54, at 348–61.

¹⁹⁸ Bertram Gold & William Salkind, *What Do ‘Top Box’ Scores Measure?*, 14 J. ADVERT. RSCH. 19, 19–23 (1974); Alin Gruber, *Purchase Intent and Purchase Probability*, 10 J. ADVERT. RSCH. 23, 23–27 (1970).

¹⁹⁹ Russell I. Haley & Peter B. Case, *Testing Thirteen Attitude Scales for Agreement and Brand Discrimination*, 43 J. MKTG. 20, 20–32 (1979); Michel Laroche & John A. Howard, *Non-Linear Relations in a Complex Model of Buyer Behavior*, 6 J. CONSUMER RSCH. 377, 377–88 (1980); see also Manohar U. Kalwani & Alvin J. Silk, *On the Reliability and Predictive Validity of Purchase Intention Measures*, 1 MKTG. SCI. 243, 243–86 (1982). Laroche and Howard employed a five-point Likert intention scale, Laroche, *supra*, at 380, while Haley and Case’s data demonstrate a threshold effect for both five-point and eleven-point Likert intention scales, Haley, *supra*, at 22–23.

strength, moreover, are more likely to impact consumer behavior. For these reasons, the party bearing the burden of persuasion on the question addressed by a survey should not be able to rely on consumers who admit to guessing or who report that they are merely somewhat likely to perceive the mark or marks at issue in the manner alleged.

Consider, for example, the likelihood-of-confusion cause of action. As its name suggests, this cause of action is probabilistic in nature. The plaintiff must establish not a possibility, but a *probability* that some appreciable proportion of relevant consumers are likely to be confused by the defendant's mark. The current *Squirt* and *Eveready* formats are blunt instruments that allow plaintiffs to meet the relevant threshold by grouping together consumers who are only somewhat likely to be confused with those that are very likely to be confused or definitely confused.

In this regard, the current survey formats take advantage of survey respondents' willingness to speculate in low-stakes, non-purchasing contexts. In our *Squirt* surveys above, of those respondents who would conventionally be classified as likely to be confused, in Group A one third were, in fact, only somewhat likely to be confused and in Group B it was one half. If these respondents are removed from the "likely to be confused" category, both groups would yield results that for most courts fail to meet the standard threshold proportion of confused consumers sufficient to trigger a finding of infringement.²⁰⁰

²⁰⁰ Some trademark scholars have long argued for a general materiality requirement in trademark confusion cases—i.e., for a requirement that plaintiffs show that confusion is likely to be material to actual purchasing decisions. *See, e.g.*, Glynn S. Lunney, Jr., *Trademark Monopolies*, 48 EMORY L.J. 367, 484 (1999) (“Even if we are willing to presume that confusion as to source is usually material, confusion as to other issues is inherently less likely to involve an issue that will actually matter to consumers when they decide which products to buy. As a result, when confusion concerns something other than source, courts should expressly require the plaintiff to establish that the confusion concerns material information, just as we do with implicit false advertising claims more generally.” (internal citations omitted)); Mark A. Lemley & Mark McKenna, *Irrelevant Confusion*, 62 STAN. L. REV. 413, 447 (2010) (arguing that certain forms of consumer confusion “should only be actionable when they can be proven material to consumers’ decisions in particular cases”); Rebecca Tushnet, *Running the Gamut from A to B: Federal Trademark and False Advertising Law*, 159 U. PA. L. REV. 1305, 1344–73 (2011) (advocating for the restoration of a materiality requirement in trademark law). Our arguments here might be understood as a kind of “systemic” materiality requirement—one which does not operate as an additional formal element of the plaintiff’s claim, but rather as a factor shaping how plaintiffs in trademark cases are required to prove their claims. We agree that plaintiffs should be required to support their claims with evidence of consumer beliefs that are reasonably certain. If plaintiffs wish to argue that consumers with weak beliefs are likely to act on those beliefs, they should be required to offer context-specific arguments—e.g., characteristics of the real-world purchasing environment that are likely to strengthen rather than dispel weakly held beliefs—and not mere assertion.

Finally, a more nuanced understanding of consumer uncertainty in the marketplace may allow for more nuanced forms of relief. With respect to injunctive relief, all of the major fields of intellectual property law have grown increasingly sensitive in recent decades to the need for courts to fashion more tailored injunctions.²⁰¹ Though trademark law has shared in this trend, it significantly lags behind patent and copyright law.

One reason for this may be that trademark courts are not provided with sufficient information about marketplace realities. Current likelihood-of-confusion surveys contribute to this problem by giving the impression that there either is or is not confusion, rather than accurately reflecting the varying degrees of confusion among consumers that our experiments suggest are likely to characterize consumer beliefs in many cases. As courts become aware of the wide diversity of consumer beliefs, they may become more comfortable with forms of injunctive relief that fall short of outright prohibitions.

With respect to damages, many circuits require that the plaintiff show actual confusion to qualify for an award of damages.²⁰² Survey evidence is routinely accepted to support this showing.²⁰³ But survey evidence that indicates primarily that survey respondents were only somewhat likely to be confused should not support a finding of actual confusion.

In sum, our experimental findings show that current survey methods deprive courts of information useful in designing effective and appropriately tailored remedies in cases where plaintiffs do prevail. Properly designed trademark surveys would provide courts with the information they need to take account of consumer belief strength when designing remedies.

For example, when a survey reveals that an appreciable number of consumers believe mistakenly that the parties' products come from the same source, but for many that belief is only weakly held, then it may be unnecessary for a court to issue an absolute prohibition on the defendant's accused mark. Disclaimers or modest changes in the defendant's mark may be sufficient to

²⁰¹ See Aurelia Hepburn-Briscoe, *Irreparable Harm in Patent, Copyright, and Trademark Cases After eBay v. Merceexchange*, 55 How. L.J. 643, 647 (2012).

²⁰² See, e.g., *Brunswick Corp. v. Spinit Reel Co.*, 832 F.2d 513, 525 (10th Cir. 1987) ("Likelihood of confusion is insufficient; to recover damages plaintiff must prove it has been damaged by actual consumer confusion or deception resulting from the violation. . . . Actual consumer confusion may be shown by direct evidence, a diversion of sales or direct testimony from the public, or by circumstantial evidence such as consumer surveys."); see also *Int'l Star Class Yacht Racing Ass'n v. Tommy Hilfiger, U.S.A., Inc.*, 80 F.3d 749, 753 (2d Cir. 1996) ("Proof of actual confusion is ordinarily required for recovery of damages for pecuniary loss sustained by the plaintiff.").

²⁰³ See *Brunswick Corp.*, 832 F.2d at 525.

disabuse consumers of weakly held mistaken beliefs. And such tailored relief may avoid imposing unneeded costs both on the plaintiff mark-owner's good faith competitors and—perhaps most importantly—on consumers who are *not* confused.

CONCLUSION

It has now been a century since courts first began to consider trademark survey evidence, and through the course of that century, the quality and utility of survey evidence has substantially deteriorated, leaving judges understandably wary of it. Perhaps because it eased the administration of surveys, perhaps because it suited their clients' interests, trademark survey experts regressed to the practice of treating consumer beliefs as simple binaries.

Meanwhile, social science understandings of consumer beliefs and behavior have advanced considerably. Marketing scholars have long since recognized that any plausible model of consumer attitude formation must take into account consumer belief strength— b_i in the expectancy-value model outlined above.²⁰⁴ Public opinion scholars have understood that some respondents are ambivalent or unwilling to admit that they do not know or do not have an opinion. As a result, attitudes expressed on surveys are not always meaningful or strongly held.²⁰⁵

This Article has sought to restore these basic insights—that consumer beliefs are not binary, but held at varying levels of strength and meaningfulness—to the administration and evaluation of trademark survey evidence. Our hope is that trademark survey evidence developed according to the protocols we have set out will better aid courts in reaching the right results in specific cases and further the trademark system's overarching goal of promoting efficient and fair competition.

While we believe that testing for consumer belief strength will significantly improve the utility of trademark survey evidence, we recognize that other simplifying assumptions built into trademark surveys remain to be overhauled. This Article has focused on the limitations of considering consumer response as binary. However, it has only scratched the surface of simplifying assumptions in trademark surveys. For example, not only is the core survey response binary, but so is sample extraction from the study's target universe—i.e., individuals in that target universe are either expected to be future purchasers, and thereby

²⁰⁴ See *supra* notes 56–58 and accompanying text.

²⁰⁵ See *supra* text accompanying notes 58–63.

2023]

CONSUMER UNCERTAINTY IN TRADEMARK LAW

547

eligible for inclusion in the sample, or they are not. But as with beliefs, propensity toward future purchasing behavior is probabilistic.²⁰⁶

In line with the above cited research on the relationship between intention and purchase, it would be reasonable to ask whether prospective purchasers should be given the same weight as established purchasers. Similar questions could be asked about the likelihood of stimulus presentation being able to mirror a marketplace context. Of course, some aspects of survey implementation need to be simplified in order to make the study practical. This Article has addressed what we believe is the most important and most easily remedied simplification at the entrance of what, in the future, might mature into a broader revision of the role of uncertainty and probabilism in trademark law.

²⁰⁶ Thanks to Irina Manta for pointing this out.